REMARKS

Reconsideration of the present application is respectfully requested. Claims 2-10, and 12-15 are pending. Claim 11 has been cancelled as belonging to a non-elected invention. The right to pursue this claim in a continuing application is reserved. No change of inventorship is necessary. Claim 1 has been cancelled and rewritten as new claims 12-15. Support for these claims is found in the claims as originally filed, and throughout the specification. In particular, see page 32, lines 15-18 regarding new claim 15. No new matter has been added.

Claims 2, 7, 9, and 10 have been amended.

Claim 2 has been amended to have proper antecedent basis. Claim 1 has been cancelled and rewritten as new claims 12-15. Claim 2 has been amended to properly depend from new claim 12. Applicant has also deleted the phrase "in sense or anti-sense orientation". Support for these amendments can be found in the specification and claims as originally filed, in particular see page 12, lines 20-23, and page 44, lines 3-18.

Claim 7 has been amended to have proper Markush format, as recommended by the Examiner.

Claim 9 has been amended to have proper antecedent basis. Claim 9 has been amended to depend from new claim 12. Applicant has changed the gene name from "RAD51" to "RAD51C" to clarify the Rad51-like sequences of the claimed invention. Applicant has also deleted the term "maize" in reference to the sequences of the invention. Support for these amendments can be found in the specification and claims as originally filed, in particular see page 4, lines 9-11.

Claim 10 has been amended to include more plants which can be used in the method of claim 9. Support for this amendment can be found in the specification and the claims as originally filed.

Applicant has amended the specification to delete references to internet hyperlinks.

Applicant has amended the abstract to recite "RAD51C" to clarify the sequences of the invention. Applicant has deleted the term "maize" in reference to the sequences of the invention. Support for these amendments can be found in the specification and claims as originally filed, in particular see page 4, lines 9-11.

Applicant includes a new Declaration executed by the inventors which includes the ZIP Code designation for each post office address.

The marked up version of these amendments is found on a separate sheet attached to this amendment and titled **"Version with Markings to Show Changes Made."** It is respectfully requested that the amendments be entered.

Election/Restriction

The Examiner has issued a restriction requirement, and has required election of either the invention of Group I (Claims 1-10) or Group II (Claim 11). Applicants hereby affirm the provisional election to prosecute the claims of Group I, with traverse, and expressly reserves the right to file a divisional application relating to and claiming the invention of Group II. No change of inventorship is required due to this election of Group I.

The Examiner further required election of one sequence for the application. The claims have been amended to remove reference to SEQ ID NOS: 3 and 5 as per the election filed September 5, 2001. The Applicants traverse the restriction requirement and therefore respectfully request reconsideration of the same. The Applicants submit the alignments referred to in the response to the restriction requirement filed September 5, 2001 in Appendix A. These alignments demonstrate the high degree of homology between SEQ ID NOS: 1, 3, and 5, as well as the encoded proteins of SEQ ID NOS: 2, 4, and 6. The polynucleotides of the present invention, as shown in Appendix A, share greater than 99% sequence identity as determined by the GAP algorithm under default parameters. Applicants believe that one search encompasses all the sequences of the invention. As the restriction to

one sequence is at the discretion of the Examiner, it is hoped the actual analyses presented will convince the Examiner to rejoin SEQ ID NOS: 1, 3 and 5 for examination of this invention. Therefore, the Applicants respectfully request withdrawal of the sequence election in this application.

Rejections under 35 U.S.C. §101:

Claims 1-10 are rejected under 35 U.S.C. §101 as not having either a credible asserted utility or a well-established utility. Claim 1 has been cancelled and rewritten as claims 12-15, the rejection will be discussed as it applies to these claims.

The Examiner asserts that isolated polynucleotides of at least 30 contiguous nucleotides encompass nucleic acid sequences encoding a mannanase (Buchert, et al., 1997, US Patent 5,66,021), a human DNA repair protein (1998, GenBank AI184177), and/or an Arabidopsis RAD57 (Rounsley et al., 1998, GenBank O22144), and that the instant specification does not teach a specific use of these nucleic acids.

Applicants respectfully disagree. Applicants do teach a specific use for the polynucleotides claimed. In claim 9, Applicants claim a method to modulate the level of Rad51C in a plant using a polynucleotide of claim 12. For example, subsequences of a nucleic acid can be used to modulate the level gene expression, see page 32, lines 23-33; and page 44, lines 3-18. Therefore, subsequences of polynucleotide sequences of the present invention do have a specific utility. Not all embodiments must have utility for the invention as a whole to have utility. Inoperable embodiments of the claimed invention do not eliminate the utility of the operable embodiments. As it is stated in the MPEP 2107 II, page 2100-25: "... as the Federal Circuit has stated, '[t]o violate [35 U.S.C.] 101 the claimed device must be totally incapable of achieving a useful result.' *Brooktree Corp. v. Advanced Micro Devices, Inc.*, 977 F.2d 1555, 1571, 24 USPQ2d 1401, 1412 (Fed. Cir. 1992)".

The nucleic acid sequence encoding a mannanase (Buchert, *et al.*, 1997, US Patent 5,66,021) shares 30 contiguous nucleotides with SEQ ID NO: 1 in the polyA tail region. Claim 12 recites "An isolated polynucleotide encoding a polypeptide with Rad51C activity", therefore the polyA tail of a nucleic acid encoding a mannanase is not encompassed by claim 12. Claim 15 claims "A polynucleotide comprising at least 50 contiguous nucleotides from a polynucleotide of SEQ ID NO: 1". Therefore, claim 15 requires at least 50 contiguous nucleotides, as such the sequence disclosed by Buchert, *et al.* is not encompassed by this claim.

Applicants do not claim polynucleotide subsequences with a given percent identity. Applicants claim at least 80% sequence identity over the full length using the GAP program, a Global Alignment Program. Applicants separately claim a polynucleotide with at least 50 **contiguous** nucleotides. Therefore, the sequence of AI184177 is not encompassed in claims 12-15.

Claim 12 recites "An isolated polynucleotide encoding a polypeptide with Rad51C activity". Therefore, nucleic acid sequences encoding a mannanase (Buchert, et al., 1997, US Patent 5,66,021), a human DNA repair protein (1998, GenBank AI184177), and/or an *Arabidopsis* RAD57 (Rounsley *et al.*, 1998, GenBank O22144) are not encompassed by claim 12.

The sequence search results provided show an **amino acid alignment** of SEQ ID NO: 2 with an *Arabidopsis* RAD57 (Rounsley *et al.*, 1998, GenBank O22144). This alignment shows 14 contiguous **amino acids** shared by the two sequences. Applicant submits evidence in Appendix B that shows that SEQ ID NO: 1 and the polynucleotide encoding an *Arabidopsis* RAD57 as disclosed by Rounsley *et al.* do not share 30 contiguous nucleotides, even though both polynucleotide sequences encode 14 contiguous amino acids. Appendix B contains two alignments. First, an alignment (FrameAlign) of the polynucleotide SEQ ID NO: 1 with the polypeptide of O22144 was done in order to identify the appropriate region of SEQ ID NO: 1. Second, a GAP alignment of the polynucleotide of SEQ ID NO: 1

and the polynucleotide encoding the polypeptide of O22144. This GAP alignment shows the two polynucleotide sequences do not share 30 or 50 contiguous nucleotides. Therefore, the sequence of O22144 was not encompassed in originally filed claim 1, or in new claims 12-15.

Claim 14 claims a polynucleotide comprising at least 100 contiguous nucleotides which selectively hybridizes, under stringent conditions and a wash in 0.1X SSC at 60°C. Support for this claim can be found in the originally filed claims and in the specification, for example page 32, lines 15-22. Applicants define "selectively hybridizes" on page 14, line 30 – page 15, line 3 of the specification. Sequences which selectively hybridize, under stringent conditions, hybridize at least 2-fold over background and to the substantial exclusion of non-target nucleic acids. It is also noted, that selectively hybridizing sequences typically have at least about 80% sequence identity with each other. "Stringent conditions" are defined and discussed on page 15, line 24 – page 17, line 11, particularly page 16, lines 6-13. The role of post-hybridization washes is also discussed. Given that the polynucleotide must be a polynucleotide of at least 100 contiguous nucleotides which selectively hybridizes, under stringent conditions, and a wash in 0.1X SSC at 60°C, nucleic acid sequences encoding a mannanase (Buchert, *et al.*, 1997, US Patent 5,66,021), a human DNA repair protein (1998, GenBank AI184177), and/or an *Arabidopsis* RAD57 (Rounsley *et al.*, 1998, GenBank O22144) are not encompassed by claim 14.

Claim 15 claims "A polynucleotide comprising at least 50 contiguous nucleotides from a polynucleotide of SEQ ID NO: 1". Nucleic acid sequences encoding a mannanase (Buchert, *et al.*, 1997, US Patent 5,66,021), a human DNA repair protein (1998, GenBank AI184177), and/or an *Arabidopsis* RAD57 (Rounsley et al., 1998, GenBank O22144) are not encompassed by claim 15.

New claims 12-15 do not encompass a subsequence of a nucleic acid encoding a mannanase (Buchert *et al.*, 1997, US Patent 5,661,021), a human DNA

repair protein (1998, GenBank AI184177), and/or an *Arabidopsis* RAD57 (Rounsley *et al.*, 1998, GenBank O22144). Subsequences of polynucleotides of the present invention have utility as discussed throughout the specification, for example see page 32, lines 23-33; and page 44, lines 3-18. Therefore the rejection under 35 U.S.C §101 of claims 1-10, as applied to claims 2-10 and 12-15, should be withdrawn.

The Examiner states claim 8 is rejected under 35 U.S.C. §101 as not having a specific or well-established utility because the claim does not require that the transgenic seed have the expression cassette of claim 2.

Applicants respectfully disagree. Claim 8 depends from claim 4, which claims a transgenic plant comprising a recombinant expression cassette of claim 2. As the transgenic plant is required to comprise a recombinant expression cassette of claim 2, this requirement carries into the dependent claim 8 directed to a transgenic seed. Applicants respectfully request the rejection of claim 8 under 35 U.S.C. §101 be withdrawn.

Applicants have properly addressed by argument and amendment the grounds for the rejection of originally filed claims 1-10 under 35 U.S.C. §101 as it would apply to pending claims 2-10, and 12-15, and respectfully request that the rejection of the claims under 35 U.S.C §101 be withdrawn.

Rejections under 35 U.S.C. §112, first paragraph – Utility:

As the Applicants have responded to the utility rejection under 35 U.S.C. §101, the concomitant rejection of claims 1-10 under 35 U.S.C. §112, first paragraph based on a lack of utility should be withdrawn and not applied to pending claims 2-10, and 12-15.

Rejections under 35 U.S.C. §112, first paragraph:

Claims 1-10 are rejected under 35 U.S.C. §112, first paragraph. The specification does not enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and/or use the invention commensurate in scope with these claims. This rejection will be discussed as it pertains to original claims 2-10, and new claims 12-15.

The Examiner states "Claims 1-10 are rejected under 35 U.S.C. §112, first paragraph, because the specification, while being enabling for nucleic acids of SEQ ID NO: 1 or that encode SEQ ID NO: 2, does not reasonably provide enablement for nucleic acids that have 80% identity to SEQ ID NO: 1, that are amplified from primers that hybridize under unspecified stringency to 'loci within' SEQ ID NO: 1, or that comprise 30 nucleotides that hybridize to SEQ ID NO: 1."

The Applicants respectfully disagree. The specification provides guidance for modification and variants of the polynucleotides and/or polypeptides of the instant invention (for example: page 7, line 12 – page 9, line 3; page 12, lines 15-31; page 25, line 24 – page 26, line 6; page 28, line 27 – page 31, line 29; page 30, lines 2-16; page 58, lines 5-29; and SEQ ID NOS: 1-6), guidance on sequence comparison and analyses (for example: page 9, lines 4-18; page 17, line 28 – page 23, line 2; page 28, line 27 – page 29, line 4; and Examples 3 and 4, pages 64-65), guidance on amplification of polynucleotides (for example: page 6, lines 1-11; page 26, line 8 – page 28, line 3; and page 37, line 25 – page 38, line 15), guidance on hybridization of polynucleotides (for example: page 14, line 30 – page 15, line 3; page 15, line 24 – page 17, line 11; page 28, lines 5-25; page 31, line 31 – page 32, line 6; page 35, line 27 – page 36, line 22; and page 37, lines 7-24) and guidance on subsequences (for example: page 32, line 9 – page 33, line 2). Thus, Applicant respectfully submits that the specification does enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and/or use the invention commensurate in scope with these claims.

The Examiner states "The instant specification, however, fails to provide guidance for which amino acids of SEQ ID NO: 2 can be altered and to which other amino acids, and which amino acids must not be changed to maintain RAD51 activity of the encoded protein."

Applicants respectfully disagree. The background discusses conserved sequences in the RAD51 family (see page 2, lines 13-19). Example 4 on pages 64-65 of the specification specifically points out conserved sequence found in SEQ ID NO: 2, including a functional domain, the Walker A box ATP-binding motif (highlighted).

At the time of filing, it was well within the capabilities of one of skill in the art to determine which amino acids could be altered. For example, methods to assay for various functions and phenotypes associated with RAD51 homologues were well-known at the time of filing, as evidenced by the documents submitted by the Applicant in an IDS filed June 23, 2000. In particular, these references disclose several assay methods including yeast two-hybrid screens (Johnson & Symington 1995; Dosanjh *et al.* 1998), DNA strand exchange (Sung 1994 and 1997; Sung and Robberson 1995), complementation (Vispe, *et al.* 1998), homologous recombination (Vispe *et al.* 1998; Xia *et al.* 1997), and gamma-irradiation (Johnson & Symington 1995). Also, at the time of filing the structure of RecA and related proteins were known, as evidenced in Appendix C, and could be used to model the structure of Rad51-like sequences and serve as guidance for allowable modifications. One of skill in the art could also use multiple sequence alignments to identify putative residues and regions which allow modification, one such multiple sequence alignment is submitted in Appendix D.

As is stated in MPEP 2164.01 "A patent need not teach, and preferably omits, what is well known in the art. *In re Buchner*, 929 F.2d 660, 661, 18 USPQ2d 1331, 1332 (Fed. Cir. 1991); *Hybritech, Inc. v. Monoclonal Antibodies, Inc.*, 802 F.2d 1367, 1384, 231 USPQ 81, 94 (Fed. Cir. 1986), *cert. denied*, 480 U.S. 947 (1987); and

Lindemann Maschinenfabrik GMBH v. American Hoist & Derrick Co., 730 F.2d 1452, 1463, 221 USPQ 481, 489 (Fed. Cir. 1984)."

The Examiner states "It cannot be predicted by one of skill in the art that nucleic acids having 80% identity to SEQ ID NO: 1, that are amplified from primers that hybridize under unspecified stringency to 'loci within' SEQ ID NO: 1, or that comprise 30 nucleotides that hybridize to SEQ ID NO: 1 will encode a protein with the same activity as SEQ ID NO: 2."

The Examiner cites Bowie *et al.*(1990, *Science*) which teaches that protein structure prediction, and ascertaining functional aspects of the protein, from sequence data is extremely complex. The Examiner also cites Lazar *et al.*, Broun *et al.*, Burgess *et al.*, and Hill *et al.*, all of which provide examples of very specific limited amino acid changes which result in elimination or alteration of the experimental protein's catalytic activity.

Applicant notes that Bowie *et al.* also teaches commonly used methods to predict tolerance of an amino acid sequence to change, observing tolerated substitutions in related sequences through evolution (*e.g.*, see Applicants Appendix D), and genetic manipulation of sequence (page 1306, paragraph bridging columns 1 and 2). Bowie *et al.* further reveals that studies using these methods reveal that proteins are highly tolerant of amino acid substitutions, with as many as one-half of all substitutions being phenotypically silent in *lac* repressor (page 1306, 1st full paragraph column 2). As is noted above, methods to assay function, as well as the structure of related proteins were available at the time of filing, coupled with the disclosures of the present application, one of skill in the art was reasonably apprised of the scope of the invention. The invention is directed to compositions of RAD51C, its activities, and methods of use, non-functional embodiments are not claimed and do not eliminate the utility of the function embodiments set forth in the claims.

The Examiner cites Reiss *et al.* (2000, *PNAS*) wherein plants transformed with RecA unexpectedly did not have an increase in gene targeting.

Applicants note that Reiss *et al.* did observe that RecA did increase the fidelity of the recombination (page 3363, left column, lines 1-2). Further, Reiss *et al.* postulate that RecA may not have increased gene targeting due to unavailability of ssDNA substrate in the *Agrobacterium*-mediated transformation method used (page 3363, left column, 1st full paragraph).

The Examiner asserts that the instant specification fails to teach how nucleic acids encoding a mannanase, or how human or *Arabidopsis* nucleic acids that do not encode RAD51, cited in the 35 U.S.C. §101 rejection, could be used to modulate the level of maize RAD51 in a plant.

As discussed in regard to the 35 U.S.C. §101 rejection, the claims of the present invention do not encompass non-RAD51-like nucleic acids or proteins.

The Examiner asserts, given the claim breadth, unpredictability, and lack of guidance, that undue experimentation would have been required by one skilled in the art to practice the invention.

The Applicants respectfully disagree. As noted above, Applicants have disclosed several sequences (SEQ ID NOS: 1-6), provided guidance regarding modifications to the sequences, methods to analyze, isolate, identify and characterize the sequences. The 3-dimensional structure of related proteins were known in the art at the time of filing, as well as methods to assay for functional RAD51 homologues. The question of experimentation is a matter of degree. The fact that some experimentation is necessary does not preclude enablement; what is required is the amount of experimentation must not be unduly extensive. *PPG Inc. v. Guardian Industries Corp.* (37 USPQ 1218, 1623, (Fed. Cir. 1996). The test is not merely quantitative, since a considerable amount of experimentation is permissible, if it is merely routine, or if the specification in question provides a reasonable amount of guidance with respect to the direction in which the experimentation should proceed to enable the determination of how to practice a desired embodiment of the invention claimed. *Ex parte Jackson*, 217 USPQ 804, 807 (1982 PTOBA).

Applicants have provided reasonable guidance such that one of skill in the art can practice the breadth of the invention as disclosed and claimed, therefore the rejection of claims 2-10 and 12-15 under 35 U.S.C. §112, first paragraph should be withdrawn.

Claims 1-10 are rejected under 35 U.S.C. §112, first paragraph, as containing subject matter that was not described in the specification in such a way as to reasonably convey to one skilled in the relevant are that the inventor(s), at the time the application was filed, had possession of the claimed invention.

The Examiner states: "Claim 1 recites no description of the function of the protein encoded by the nucleic acid and the plant claims recite no phenotype."

Claim 1 has been cancelled and rewritten as new claims 12-15. This rejection will be addressed as it may be applied to these claims.

Claim 12 recites the function of the protein in the preamble, "An isolated polynucleotide encoding a polypeptide with Rad51C activity". Therefore the rejection to claim 1 should not be applied to claim 12.

Applicant notes that the phenotype of the transgenic plants claimed will depend on the components and orientations of recombinant expression cassette constructed. For example, a transgenic plant in which a developmental-specific, pollen-specific promoter is used to drive transcription of SEQ ID NO: 1 in the antisense orientation will have a different phenotype than a transgenic plant in which SEQ ID NO: 1 is operably linked in the sense orientation to a strong, constitutive promoter. Applicant need not recite the plant phenotype in the claims, the metes and bounds of claim 4 and dependent claims are clear, the transgenic plant comprises an isolated polynucleotide of claim 12 in a recombinant expression cassette.

The Examiner cites Dosanjh *et al.* 1998 (*Nucl. Acids Res.* 26:1179-1184; 1183 right column, paragraph 2) to support the assertion that different RAD51 proteins appear to have different functions within a cell.

Dosanjh *et al.* report the discovery of another member of the RAD51 gene family. The different members of the RAD51 gene family are components that function together in the recombinational repair of DNA (see page 1170, first paragraph, bridges columns 1 and 2).

The Examiner cites *University of California v. Eli Lilly*, 119 F.3d 1559, 43 USPQ d2 1398 (Fed Cir. 1997) which states "...there is no further information in the patent pertaining to that cDNA's relevant structural or physical characteristics; in other words, it thus does not describe human insulin cDNA...". The Examiner also points out, on page 1046: "A definition by function, as we have previously indicated, does not suffice to define the genus because it is only an indication of what the gene does, not what it is."

The Examiner also cites *Amgen Inc. v. Chugai Pharmaceutical Co. Ltd.*, 18 USPQ 2d 1016 at 1021 (Fed Cir. 1991). The Examiner directs Applicants' attention to page 1021: "Conception does not occur unless one has a mental picture of the structure of the chemical or is able to define it by its method of preparation, its physical or chemical properties, or whatever characteristics sufficiently distinguish it."

It should be noted that in both of the above cases, the claim language focused on the biological properties of the claimed sequences. Applicants respectfully submit that the present invention is not defined solely by its biological property but is defined by structural features such as percent identity and hybridization fidelity. These structural features are readily understood by those practicing the art and are fully supported in the specification. For example, claim 12 claims polynucleotides having at least 80% sequence identity to the polynucleotides of SEQ ID NO: 1; wherein the percent sequence identity is based on the entire coding regions and is determined by the GAP program under default parameters. While SEQ ID NO: 1 is clearly defined in the sequence listing, those polynucleotides having at least 80% sequence identity to the polynucleotides of SEQ ID NO: 1 are

clearly defined in the instant specification. The definition of sequence identity is taught on page 21, line 29 – page 22, line 5 of the specification. The description of sequence similarity, methods for aligning sequences, and a description of the GAP program used to determine the percentage of sequence identity can be found on page 18, line 14 – page 23, line 2 of the specification. Further, methods of making the invention are also clearly taught. See, for example, pages 34 - 37 where library synthesis (page 34, line 6 – page 35, line 25; and page 36, line 24 – page 37, line 5), screening of DNA libraries (page 37, line 7 – page 38, line 15), amplification of polynucleotides (page 6, lines -11; page 26, line 8 - page 27, line 29; and page 37, line 25 – page 38, line 15), and synthetic preparation of polynucleotides (page 38, lines 17-31) are taught. See page 15, line 24 – page 17, line 11; and page 37, lines 8-24 for a description of hybridization conditions.

In *Amgen v. Chugai*, the Federal Circuit concluded that the patent specification was insufficient to enable one of ordinary skill in the art to make and use the invention claimed in claim 7 of the '008 patent without undue experimentation. As stated on page 1027, however, "it is not necessary that a patent applicant test all the embodiments of his invention, In re Angstadt, 537 F.2d 498, 502, 190 USPQ 214, 218 (CCPA 1976); what is necessary is that he provide a disclosure sufficient to enable one skilled in the art to carry out the invention commensurate with the scope of his claims. For DNA sequences, that means disclosing how to make and use enough sequences to justify grant of the claims sought." Applicants respectfully submit, that has been done in the instant specification. The present invention discloses how to make and use the sequences of the invention, as discussed in the paragraph above.

The question of experimentation is a matter of degree. The fact that some experimentation is necessary does not preclude enablement; what is required is the amount of experimentation must not be unduly extensive. *PPG Inc. v. Guardian Industries Corp.* (37 USPQ 1218, 1623, (Fed. Cir. 1996).

The present specification provides reasonable guidance with respect to the direction in which the experimentation should proceed by providing sequences, methods, citations and examples sufficient to practice the scope of the claims. While the methods require selection of transformed plants exhibiting the desired traits/phenotype, the selection is routine and would not require undue experimentation. No matter how much detail is provided, one will have to select for the desired phenotype.

The test is not merely quantitative, since a considerable amount of experimentation is permissible, if it is merely routine, or if the specification in question provides a reasonable amount of guidance with respect to the direction in which the experimentation should proceed to enable the determination of how to practice a desired embodiment of the invention claimed. *Ex parte Jackson*, 217 USPQ 804, 807 (1982 PTOBA).

With the guidance provided in the present specification, one skilled in the art can readily practice the claimed invention. Therefore, it is respectfully requested that the rejection of claims 1-10 under 35 U.S.C. §112, first paragraph be withdrawn and not applied to pending claims 2-10, and 12-15.

Rejections under 35 U.S.C. §112, second paragraph:

Claims 1-10 have been rejected under 35 U.S.C. § 112, second paragraph as being indefinite. Claim 1 has been cancelled and rewritten as claims 12-15. This rejection will be addressed as it applies to pending claims 2-10, and 12-15.

The Examiner asserts claim 1 is indefinite in its recitation of "GAP algorithm". This rejection will be discussed as it applies to new claim 12.

New claim 12 recites "GAP program", as recommended by the Examiner, therefore the rejection should not be applied to claim 12.

The Examiner asserts claim 1, parts (c) and (d) are indefinite in the recitation of "stringent hybridization conditions" and "selectively hybridize(s)". This rejection

will be discussed as it applies to new claims 13 and 14. The Examiner further asserts that claim 1, part (d) is indefinite for not indicating the length of wash time. This rejection will be discussed as it applies to claim 14.

Hybridization is a common technique to those of skill in the art, as is illustrated by the availability of commercial kits as well as several standard references including Sambrook *et al.* (1989) *Molecular Cloning – A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Press; Ausubel *et al.*, Eds. (1994) *Current Protocols in Molecular Biology*, Greene Publishing Assoc., Inc. and John Wiley and Sons, Inc.; and Berger and Kimmel, *Guide to Molecular Cloning Techniques, Methods in Enzymology*, Vol. 152, Academic Press.

As it is defined in the specification on page 14, line 30 – page 15, line 3, "selectively hybridize(s)":

> 'includes reference to hybridization, *under stringent hybridization conditions*, of a nucleic acid sequence to a specified nucleic acid target sequence to a detectably greater degree (e.g., *at least 2-fold over background*) than its hybridization to non-target nucleic acid sequences and to the substantial exclusion of non-target nucleic acids.'

"Selectively hybridizes", as defined in the specification, indicates selective hybridization is at least 2-fold over background as compared to a non-target nucleic acid under stringent hybridization conditions.

"Stringent hybridization conditions" are discussed extensively on pages 15-17 of the specification. Stringent conditions are those at which the probe will **selectively hybridize** to its target at least 2-fold over background as compared to non-target nucleic acids. It is noted that these conditions will be sequence dependent, but guidance on conditions is given on pages 16-17.

The role of wash conditions is discussed on pages 16-17. While temperature and ionic strength are viewed as important factors, the time of the wash, in general is not. Claim 14 defines the critical wash parameters of ionic strength and temperature as 0.1X SSC at 60°C.

The test for definiteness is whether one skilled in the art would understand the bounds of the claim when read in light of the specification. The Examiner is reminded, to satisfy the requirements of §112, second paragraph, the claims need only "reasonably apprise those skilled in the art" as to their scope and be "as precise as the subject matter permits". *Hybritech Inc. v. Monoclonal Antibodies, Inc.*, 802 F.2d 1367, 231 USPQ 81 (Fed. Cir. 1986), cert. denied, 480 U.S. 947 (1987). The language of claims 13 and 14 is not ambiguous when read in light of the specification.

Accordingly, claims 13 and 14 fulfills the requirements of 35 U.S.C. §112, second paragraph, and the Examiner is respectfully requested to withdraw the rejection of claim 1, parts (c) and (d) and not apply the rejection to newly submitted claims 13 and 14.

The Examiner asserts that claim 7 is not written in proper Markush format, as it has improper punctuation after the phrase "selected from the group consisting of".

The Applicant has amended claim 7 to remove the colon after the phrase "selected from the group consisting of", as recommended by the Examiner. Claim 7 is now in proper form and the rejection under 35 U.S.C. §112, second paragraph should not be applied to the amended claim.

Applicant has addressed the rejections under 35 U.S.C. §112, second paragraph by proper amendments and arguments. Claims 2-15 are in proper form, therefore Applicant respectfully requests the rejections under 35 U.S.C. §112, second paragraph be withdrawn.

Rejections under 35 U.S.C. § 102:

Claims 1-3 have been rejected under 35 U.S.C. § 102(b) as being anticipated by Buchert *et al.* (1997, US Patent 5,661,021).

The Examiner asserts "Buchert *et al.* teach a polynucleotide comprising at least 30 contiguous nucleotides of SEQ ID NO: 1. This nucleic acid was in an expression cassette and expressed in yeast cells."

Claim 1 has been cancelled and rewritten as claims 12-15. The rejection will be addressed as it may be applied to new claims 12-15, as well as claims 2 and 3.

The Applicants respectfully traverse the rejection under 35 U.S.C. § 102(b). As it is stated in the MPEP 2131 page 2100-54 "To anticipate a claim, the reference must teach every element of the claim. 'A claim is anticipated only if each and every element as set forth in the claim is found, either expressly or inherently described, in a single prior art reference.'"

Buchert *et al.* teach mannanase enzymes and uses thereof. Claims 12 and 13 claim isolated polynucleotides which encode a polypeptide with RAD51C activity. Claim 14 claims an isolated polynucleotide comprising at least 100 contiguous nucleotides which selectively hybridizes, under stringent conditions, to SEQ ID NO: 1. Claim 15 claims and isolated polynucleotide comprising at least 50 contiguous nucleotides from a polynucleotide of SEQ ID NO: 1. Buchert *et al.* does not disclose polynucleotides which encode a polypeptide with RAD51C activity, polynucleotides of at least 100 contiguous nucleotides which selectively hybridize to SEQ ID NO: 1 of the instant invention, or polynucleotides of at least 50 contiguous polynucleotides of SEQ ID NO: 1, therefore Buchert *et al.* does not anticipate claims 12-15, or claims 2 and 3.

Claims 1-3 have been rejected under 35 U.S.C. § 102(a) as being anticipated by NCI-CGAP (1998, GenBank Accession No. Ai184177).

The Examiner asserts "NCI-CGAP teaches a nucleic acid that comprises 40 nucleotides with more than 80% sequence identity to SEQ ID NO: 1. This nucleic acid would selectively hybridize to SEQ ID NO: 1. This nucleic acid would be in an expression cassette like that of a pUC or similar vector and this would be in a host cell for purposes of molecular biological manipulation.

Claim 1 has been cancelled and rewritten as claims 12-15. The rejection will be addressed as it may be applied to new claims 12-15, as well as claims 2 and 3.

Applicants respectfully disagree, claims 2 and 3 depend from new claim 12. New claim 12 does not encompass polynucleotides which hybridize to SEQ ID NO: 1. Claim 14 claims an isolated polynucleotide comprising at least 100 contiguous nucleotides which selectively hybridizes, under stringent conditions, to SEQ ID NO: 1. NCI-CGAP does not teach a nucleic acid of 100 contiguous nucleotides which would selectively hybridize under the claimed conditions, therefore NCI-CGAP does not anticipate claims 12-14, or claims 2-3.

Claims 1-3 have been rejected under 35 U.S.C. § 102(b) as being anticipated by Rounsely *et al.* (1998, GenBank Accession No. O22144).

The Examiner asserts "Rounsley *et al.* teach a nucleic acid that comprises 42 contiguous nucleotides that encodes SEQ ID NO: 2. This nucleic acid would be in an expression cassette like that of a pUC or similar vector and this would be in a host cell for purposes of molecular biological manipulation."

Claim 1 has been cancelled and rewritten as claims 12-15. The rejection will be addressed as it may be applied to new claims 12-15, as well as claims 2 and 3.

The sequence search results provided show an amino acid alignment of SEQ ID NO: 2 with an *Arabidopsis* RAD57 (Rounsley *et al.*, 1998, GenBank O22144). This alignment shows 14 contiguous amino acids shared by the two sequences. Applicant submits evidence in Appendix B that shows that SEQ ID NO: 1 and the polynucleotide encoding an *Arabidopsis* RAD57 as disclosed by Rounsley *et al.* do not share 30 contiguous nucleotides, even though both polynucleotide sequences encode 14 contiguous amino acids. This GAP alignment shows the two polynucleotide sequences do not share 30 contiguous nucleotides. Further, Rounsley *et al.* do not disclose a polynucleotide, or any vectors or host cells comprising a polynucleotide, those features are merely inferred by the Examiner.

Therefore, the sequence of O22144 does not anticipate originally filed claims 1-3, or new claims 12-15.

Applicants respectfully request that the rejections to claims 1-3 under 35 U.S.C. § 102(a) and 102(b) should be withdrawn and not applied to new claims 12-15.

Rejections under 35 U.S.C. § 103:

Claims 1-4, 6, and 8-9 have been rejected under 35 U.S.C. § 103(a) as being unpatentable over Reiss *et al.* (1996, *Proc. Natl. Acad. Sci.* 93:3094-3098) in view of Rounsley *et al.* (*supra*).

Claim 1 has been cancelled and rewritten as claims 12-15. The rejection will be addressed as it may be applied to new claims 2-4, 6, 8-9, and 12-15.

New claims 12-15, disclose novel and non-obvious RAD51-like sequences that are not disclosed in Reiss *et al.* or Rounsley *et al.* separately, or in combination. The disclosure of Rounsley *et al.* is discussed above and illustrated in Appendix B. The combination of Reiss *et al.* and Rounsley *et al.* does not yield the polynucleotides, methods, or compositions of the present invention. It is respectfully requested that the rejection of claims 2-4, 6, and 8-9 under 35 U.S.C. § 103(a) be withdrawn, and that the rejection not be applied to new claims 12-15.

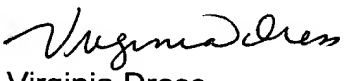
Claims 5, 7, and 10 have been rejected under 35 U.S.C. § 103(a) as being unpatentable over Reiss *et al.* (*supra*) in view of Rounsley *et al.* (*supra*), and further in view of Gordon-Kamm *et al.* (1990, *Plant Cell* 2:603-618).

Claims 5, 7 and 10 depend from new claim 12, which discloses novel and non-obvious RAD51-like sequences that are not disclosed in Reiss *et al.* or Rounsley *et al.* or Gordon-Kamm *et al.* separately, or in combination. The combination of Reiss *et al.*, Rounsley *et al.* and Gordon-Kamm *et al.* does not yield the compositions of claims 5, 7, and 10. Therefore it is respectfully requested that the rejection of claims 5, 7, and 10 under 35 U.S.C. § 103(a) be withdrawn.

CONCLUSION

In light of the foregoing remarks and amendments, withdrawal of the outstanding rejections and allowance of all of the remaining claims is respectfully requested. Applicants believe that the claims are in condition for allowance. The Examiner is invited to telephone the Applicant in order to expedite prosecution of the application.

Respectfully submitted,



Virginia Dress
Agent for Applicant(s)
Registration No. 48,243

PIONEER HI-BRED INTERNATIONAL, INC.
Corporate Intellectual Property
7100 N.W. 62nd Avenue
P.O. Box 1000
Johnston, Iowa 50131-1000
Phone: (515) 270-4192
Facsimile: (515) 334-6883

VERSION WITH MARKINGS TO SHOW CHANGES MADE

The Applicants have used underlining to denote additions to the original text and square brackets [ ] to denote deletions of the original text.

In the Title:

The title found on the cover page has been amended as follows:

[A Novel Maize] Rad51-Like [Gene] Orthologues and Uses Thereof

In the Specification:

Paragraph beginning at line 3 of page 19 has been amended as follows:

Software for performing BLAST analyses is publicly available, e.g., through the National Center for Biotechnology Information [(http://www.ncbi.nlm.nih.gov/)]. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold. These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of

the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff (1989) Proc. Natl. Acad. Sci. USA 89:10915).

Paragraph beginning at line 8 of page 64 has been amended as follows:

Gene identities were determined by conducting BLAST (Basic Local Alignment Search Tool; Altschul, S. F., et al., (1990) J. Mol. Biol. 215:403-410[; see also www.ncbi.nlm.nih.gov/BLAST/]) searches under default parameters for similarity to sequences contained in the BLAST "nr" database (comprising all non-redundant GenBank CDS translations, sequences derived from the 3-dimensional structure Brookhaven Protein Data Bank, the last major release of the SWISS-PROT protein sequence database, EMBL, and DDBJ databases). The cDNA sequences were analyzed for similarity to all publicly available DNA sequences contained in the "nr" database using the BLASTN algorithm. The DNA sequences were translated in all reading frames and compared for similarity to all publicly available protein sequences contained in the "nr" database using the BLASTX algorithm (Gish, W. and States, D. J. Nature Genetics 3:266-272 (1993)) provided by the NCBI. In some cases, the sequencing data from two or more clones containing overlapping segments of DNA were used to construct contiguous DNA sequences.

In the Abstract:

The Abstract beginning at line 1 of page 68 has been amended as follows:

ABSTRACT OF THE DISCLOSURE

The invention provides isolated [maize] RAD51C nucleic acids and their encoded proteins. The present invention provides methods and compositions relating to altering [maize] RAD51C levels in plants. The invention further provides recombinant expression cassettes, host cells, transgenic plants, and antibody compositions.

In the Claims:

Claims 1 and 11 have been cancelled without prejudice.

Claims 2, 7, 9 and 10 have been amended as follows:

2. (Amended) A recombinant expression cassette comprising a member of claim [1] 12 operably linked[, in sense or anti-sense orientation,] to a promoter.

7. (Amended) The transgenic plant of claim 4, wherein said plant is selected from the group consisting of[:] maize, soybean, sunflower, sorghum, canola, wheat, alfalfa, cotton, rice, barley, and millet.

9. (Amended) A method of modulating the level of [maize] RAD51C in a plant, comprising:

(a) introducing into a plant cell a recombinant expression cassette comprising a [maize RAD51] polynucleotide of claim [1] 12 operably linked to a promoter;

(b) culturing the plant cell under plant cell growing conditions;

(c) regenerating a whole plant which possesses the transformed genotype; and

(d) inducing expression of said polynucleotide for a time sufficient to modulate the level of [maize] RAD51C in said plant.

10. (Amended) The method of claim 9, wherein the plant is selected from the group consisting of maize, soybean, sunflower, sorghum, canola, wheat, alfalfa, cotton, rice, barley, and millet.

New claims 12-15 have been added as follows:

12. An isolated polynucleotide encoding a polypeptide with Rad51C activity comprising a member selected from the group consisting of:

(a) a polynucleotide having at least 80% sequence identity over the entire length of the reference sequence, as determined by the GAP program under default parameters, to a polynucleotide of SEQ ID NO: 1;

(b) a polynucleotide encoding a polypeptide of SEQ ID NO: 2;

(c) a polynucleotide of SEQ ID NO: 1;

(d) a polynucleotide which is fully complementary to a polynucleotide of (a), (b), or (c).

13. An isolated polynucleotide amplified from a *Zea mays* nucleic acid library using primers which selectively hybridize, under stringent hybridization

conditions, to loci within a polynucleotide of SEQ ID NO: 1, wherein the polynucleotide encodes a polypeptide with Rad51C activity.

14. An isolated polynucleotide comprising at least 100 contiguous nucleotides which selectively hybridizes, under stringent hybridization conditions and a wash in 0.1X SSC at 60°C, to a polynucleotide of SEQ ID NO: 1.

15. An isolated polynucleotide comprising at least 50 contiguous nucleotides from a polynucleotide of claim 12.

APPENDIX A

Gap Results

GAP of: 1107sid3 check: 3152 from: 1 to: 1456

WPDEF Case 1107 SEQ ID NO: 3
Case 1107 Rad51-like sequences SEQ ID NO: 3 from SEQ LISTING

to: 1107sid5 check: 9084 from: 1 to: 1333

WPDEF Case 1107 SEQ ID NO: 5
Case 1107 SEQ ID NO: 5 from SEQ LISTING. Rad51-like sequences

Symbol comparison table: nwsgapdna.cmp CompCheck: 8760

Gap Weight:	50	Average Match:	10.000
Length Weight:	3	Average Mismatch:	0.000
Quality:	12858	Length:	1470
Ratio:	9.646	Gaps:	4
Percent Similarity:	99.318	Percent Identity:	99.242

Match display thresholds for the alignment(s):
| = IDENTITY
: = 5
. = 1

1107sid3 x 1107sid5 August 28, 2001 16:14 ..

```
      1 cgacgtaagcggctgcgtggcgccaccgacggaggctacgagcggttgtg 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 cgacgtaagcggctgcgtggcgccaccgacggaggctacgagcggttgtg 50

     51 gaggcagatatgagaggtggaggtggctacaacgggtcggcggctgtgag 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 gaggcagatatgagaggtggaggtggctacaacgggtcggcggctgtgag 100

    101 atactgaaatccgcactgcagttctcttcttcccccaatcagtaccacct 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    101 atactgaaatccgcactgcagttctcttcttcccccaatcagtaccacct 150

    151 ctccaagtggcaatcaccatggga...caatctggctctagaaatggacc 197
        |||||||||||||||||||||||   |||||||||||||||||||||||
    151 ctccaagtggcaatcaccatgggagatcaatctggctctagaaatggacc 200

    198 acaacagaagtacgtttcaggagcccagaatgcctgggatatgttctctg 247
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    201 acaacagaagtacgtttcaggagcccagaatgcctgggatatgttctctg 250

    248 atgagctgtcacagaaacacatcactactggttctggtgacctcaatgac 297
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    251 atgagctgtcacagaaacacatcactactggttctggtgacctcaatgac 300

    298 atacttggtggcgggattcactgcaaagaagttactgagatcggtggcgt 347
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    301 atacttggtggcgggattcactgcaaagaagttactgagatcggtggcgt 350

    348 cccaggggttggtaaaactcaactggggattcaactagcaatcaatgtac 397
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    351 cccaggggttggtaaaactcaactggggattcaactagcaatcaatgtac 400

    398 aaatcccagtggaatgtggtggccttggtgggaaagcagtttatat.... 443
```

```
     ||||||||||||||||||||||||||||||||||||||||||||||||||
401  aaatcccagtggaatgtggtggccttggtgggaaagcagtttatatagat 450
          .         .         .         .         .
444  ..agagggcagtttcatggttgaacgtgtctaccagattgctgaagggtg 491
       ||||||||||||||||||||||||||||||||||||||||||||||||||
451  acagagggcagtttcatggttgaacgtgtctaccagattgctgaagggtg 500
          .         .         .         .         .
492  tattagggacatactggagcactttccgcacagccatgagaagtcctctt 541
     ||||||||||||||||||||||||||||||||||||||||||||||||||
501  tattagggacatactggagcactttccgcacagccatgagaagtcctctt 550
          .         .         .         .         .
542  ctgtccaaaaacaattacagcctgagcgtttcctggcggatatctattac 591
     ||||||||||||||||||||||||||||||||||||||||||||||||||
551  ctgtccaaaaacaattacagcctgagcgtttcctggcggatatctattac 600
          .         .         .         .         .
592  ttccggatatgcagttacaccgaacaaattgcagtcataaactacatgga 641
     ||||||||||||||||||||||||||||||||||||||||||||||||||
601  ttccggatatgcagttacaccgaacaaattgcagtcataaactacatgga 650
          .         .         .         .         .
642  gaagttcctcagagagcataaagatgtgcgtatagttattattgatagtg 691
     ||||||||||||||||||||||||||||||||||||||||||||||||||
651  gaagttcctcagagagcataaagatgtgcgtatagttattattgatagtg 700
          .         .         .         .         .
692  ttactttccactttcgacaagattttgaagatctggcactgaggaccaga 741
     ||||||||||||||||||||||||||||||||||||||||||||||||||
701  ttactttccactttcgacaagattttgaagatctggcactgaggaccaga 750
          .         .         .         .         .
742  gtgctaagtggattatcattgaagttaatgaagattgcaaagacatataa 791
     ||||||||||||||||||||||||||||||||||||||||||||||||||
751  gtgctaagtggattatcattgaagttaatgaagattgcaaagacatataa 800
          .         .         .         .         .
792  cttggcagttgtcttgttgaaccaagtcactactaaatttacagaagggt 841
     ||||||||||||||||||||||||||||||||||||||||||||||||||
801  cttggcagttgtcttgttgaaccaagtcactactaaatttacagaagggt 850
          .         .         .         .         .
842  catttcaattgactcttgctctaggtgacagctggtcccactcatgcacg 891
     ||||||||||||||||||||||||||||||||||||||||||||||||||
851  catttcaattgactcttgctctaggtgacagctggtcccactcatgcacg 900
          .         .         .         .         .
892  aaccggttgattctgcactggaatgggaacgaacgatacgcacatcttga 941
     ||||||||||||||||||||||||||||||||||||||||||||||||||
901  aaccggttgattctgcactggaatgggaacgaacgatacgcacatcttga 950
          .         .         .         .         .
942  taagtctccttcacttccagtagcctcagccccgtatgcagtgacaggca 991
     |||||||||||||||||||||||||| ||||||||||||||||||||||
951  taagtctccttcacttccagtagcctcagcaccgtatgcagtgacaggca 1000
          .         .         .         .         .
992  aagggattagagatg.tgtgagctcaaaccacaagcgagcccgagtaacg 1040
     |||||||||||||| ||||||||||||||||||||||||||||||||||
1001 aagggattagagatgctgtgagctcaaaccacaagcgagcccgagtaacg 1050
          .         .         .         .         .
1041 tagcattcttggtgtcaagcacttgtatgtccactacgctcctgcagctt 1090
     |||||||||||||||||||||||||||||||||||||||||||| ||||
1051 tagcattcttggtgtcaagcacttgtatgtccactacgctcctgctgctt 1100
          .         .         .         .         .
1091 tcttcgccatggatcttttggactagtgaggtgagactggagaatagtac 1140
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1101 tcttcgccatggatcttttggactagtgaggtgagactggagaatagtac 1150
          .         .         .         .         .
1141 cat....ttgattctcagttgctttgtgccgttggctaccaaccaacctt 1186
     |||    ||||||||||||||||||||||||||||||||||||||||||
1151 cattttgttgattctcagttgctttgtgccgttggctaccaaccaacctt 1200
          .         .         .         .         .
```

```
1187 aagagagaagtaaatacaacagaacaggctaatatagtgttttgtatctg 1236
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1201 aagagagaagtaaatacaacagaacaggctaatatagtgttttgtatctg 1250
               .         .         .         .         .
1237 aacatctggcccatcgtacattcagtaaagcctataatagcgggcatata 1286
     |||||||||:||||||||||||||||||||||||||||||||||||||||
1251 aacatctggsccatcgtacattcagtaaagcctataatagcgggcatata 1300
               .         .         .         .         .
1287 tgtgcttctctgatcaccgatcagcaaaaaaaaaaaaaaaaaaaaaaaa 1336
     ||||||||||||||||   |   |  ||||||||
1301 tgtgcttctctgatcaaaaaaaaaaaaaaaaa..................  1333
                              .
                              .
                              .
```

Input Sequence: 1107sid3

```
!!NA_SEQUENCE 1.0
WPDEF Case 1107 SEQ ID NO: 3
Case 1107 Rad51-like sequences SEQ ID NO: 3 from SEQ LISTING
1107sid3  Length: 1456  August 28, 2001 15:56  Type: N
Check: 3152  ..

       1  cgacgtaagc ggctgcgtgg cgccaccgac ggaggctacg
agcggttgtg

      51  gaggcagata tgagaggtgg aggtggctac aacgggtcgg
```

View Sequence

Input Sequence: 1107sid5

```
!!NA_SEQUENCE 1.0
WPDEF Case 1107 SEQ ID NO: 5
Case 1107 SEQ ID NO: 5 from SEQ LISTING.  Rad51-like
sequences
1107sid5  Length: 1333  August 28, 2001 16:00  Type: N
Check: 9084  ..

       1  cgacgtaagc ggctgcgtgg cgccaccgac ggaggctacg
agcggttgtg
```

View Sequence

Gap Results

GAP of: 1107sid4 check: 3715 from: 1 to: 281

WPDEF Case 1107 SEQ ID NO: 4
Case 1107 SEQ ID NO: 4 from SEQ LISTING. Rad 51-like sequences. Protein

to: 1107sid6 check: 4041 from: 1 to: 294

WPDEF Case 1107 SEQ ID NO: 6
Case 1107 SEQ ID NO: 6 from SEQ LISTING. Rad51-like sequences. Protein.

Symbol comparison table: blosum62.cmp CompCheck: 6430
BLOSUM62 amino acid substitution matrix.
Reference: Henikoff, S. and Henikoff, J. G. (1992). Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. USA 89: 10915-10919.

Gap Weight:	8	Average Match:	2.912
Length Weight:	2	Average Mismatch:	-2.003
Quality:	1449	Length:	294
Ratio:	5.157	Gaps:	1
Percent Similarity:	99.644	Percent Identity:	99.644

Match display thresholds for the alignment(s):
| = IDENTITY
: = 2
. = 1

1107sid4 x 1107sid6 August 28, 2001 16:17 ..

```
                .         .         .         .         .
      1 MGDQSGSRNGPQQKYVSGAQNAWDMFSDELSQKHITTGSGDLNDILGGGI 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 MGDQSGSRNGPQQKYVSGAQNAWDMFSDELSQKHITTGSGDLNDILGGGI 50
                .         .         .         .         .
     51 HCKEVTEIGGVPGVGKTQLGIQLAINVQIPVECGGLGGKAVYI..EGSFM 98
        |||||||||||||||||||||||||||||||||||||||||||  |||||
     51 HCKEVTEIGGVPGVGKTQLGIQLAINVQIPVECGGLGGKAVYIDTEGSFM 100
                .         .         .         .         .
     99 VERVYQIAEGCIRDILEHFPHSHEKSSSVQKQLQPERFLADIYYFRICSY 148
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    101 VERVYQIAEGCIRDILEHFPHSHEKSSSVQKQLQPERFLADIYYFRICSY 150
                .         .         .         .         .
    149 TEQIAVINYMEKFLREHKDVRIVIIDSVTFHFRQDFEDLALRTRVLSGLS 198
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    151 TEQIAVINYMEKFLREHKDVRIVIIDSVTFHFRQDFEDLALRTRVLSGLS 200
                .         .         .         .         .
    199 LKLMKIAKTYNLAVVLLNQVTTKFTEGSFQLTLALGDSWSHSCTNRLILH 248
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    201 LKLMKIAKTYNLAVVLLNQVTTKFTEGSFQLTLALGDSWSHSCTNRLILH 250
                .         .         .         .
    249 WNGNERYAHLDKSPSLPVASAPYAVTGKGIRDV..... 281
        |||||||||||||||||||||||||||||
    251 WNGNERYAHLDKSPSLPVASAPYAVTGKGIRDAVSSNHKRARVT 294
```

Input Sequence: 1107sid4

```
!!AA_SEQUENCE 1.0
WPDEF Case 1107 SEQ ID NO: 4
Case 1107 SEQ ID NO: 4 from SEQ LISTING.  Rad 51-like
sequences. Protein
 1107sid4  Length: 281  August 28, 2001 16:06  Type: P
Check: 3715  ..

       1  MGDQSGSRNG PQQKYVSGAQ NAWDMFSDEL SQKHITTGSG
DLNDILGGGI
```

View Sequence

Input Sequence: 1107sid6

```
!!AA_SEQUENCE 1.0
WPDEF Case 1107 SEQ ID NO: 6
Case 1107 SEQ ID NO: 6 from SEQ LISTING. Rad51-like
sequences. Protein.
 1107sid6  Length: 294  August 28, 2001 16:07  Type: P
Check: 4041  ..

       1  MGDQSGSRNG PQQKYVSGAQ NAWDMFSDEL SQKHITTGSG
DLNDILGGGI
```

View Sequence

Gap Results

GAP of: 1107sid1 check: 4817 from: 1 to: 1474

WPDEF Case 1107 SEQ ID NO: 1
Case 1107 Rad51-like sequences. From SEQ LISTING.

to: 1107sid3 check: 3152 from: 1 to: 1456

WPDEF Case 1107 SEQ ID NO: 3
Case 1107 Rad51-like sequences SEQ ID NO: 3 from SEQ LISTING

Symbol comparison table: nwsgapdna.cmp CompCheck: 8760

Gap Weight:	50	Average Match:	10.000
Length Weight:	3	Average Mismatch:	0.000
Quality:	12801	Length:	1611
Ratio:	8.792	Gaps:	5
Percent Similarity:	99.924	Percent Identity:	99.924

Match display thresholds for the alignment(s):
| = IDENTITY
: = 5
. = 1

1107sid1 x 1107sid3 August 28, 2001 16:13 ..

```
                                        .
                                        .
                                        .
                .           .           .           .           .
  101 acggcgcggcgcgactccccctaagcgacagcggcggcgtcgacgtaag 150
    1 ........................................|||||||||
                .           .           .           .           .
  151 cggctgcgtggcgccaccgacggaggctacgagcggttgtggaggcagat 200
      ||||||||||||||||||||||||||||||||||||||||||||||||||
   10 cggctgcgtggcgccaccgacggaggctacgagcggttgtggaggcagat 59
                .           .           .           .           .
  201 atgagaggtggaggtggctacaacgggtcggcggctgtgagatactgaaa 250
      ||||||||||||||||||||||||||||||||||||||||||||||||||
   60 atgagaggtggaggtggctacaacgggtcggcggctgtgagatactgaaa 109
                .           .           .           .           .
  251 tccgcactgcagttctcttcttcccccaatcagtaccacctctccaagtg 300
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  110 tccgcactgcagttctcttcttcccccaatcagtaccacctctccaagtg 159
                .           .           .           .           .
  301 gcaatcaccatgggagatcaatctggctctagaaatggaccacaacagaa 350
      ||||||||||||||||   ||||||||||||||||||||||||||||||||
  160 gcaatcaccatggga...caatctggctctagaaatggaccacaacagaa 206
                .           .           .           .           .
  351 gtacgtttcaggagcccagaatgcctgggatatgttctctgatgagctgt 400
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  207 gtacgtttcaggagcccagaatgcctgggatatgttctctgatgagctgt 256
                .           .           .           .           .
  401 cacagaaacacatcactactggttctggtgacctcaatgacatacttggt 450
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  257 cacagaaacacatcactactggttctggtgacctcaatgacatacttggt 306
                .           .           .           .           .
  451 ggcgggattcactgcaaagaagttactgagatcggtggcgtcccaggggt 500
      ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
 307 ggcgggattcactgcaaagaagttactgagatcggtggcgtcccaggggt 356
                .         .         .         .         .
 501 tggtaaaactcaactggggattcaactagcaatcaatgtacaaatcccag 550
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 357 tggtaaaactcaactggggattcaactagcaatcaatgtacaaatcccag 406
                .         .         .         .         .
 551 tggaatgtggtggccttggtgggaaagcagtttatatagatacagagggc 600
     ||||||||||||||||||||||||||||||||||||      |||||||
 407 tggaatgtggtggccttggtgggaaagcagtttatat......agagggc 450
                .         .         .         .         .
 601 agtttcatggttgaacgtgtctaccagattgctgaagggtgtattaggga 650
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 451 agtttcatggttgaacgtgtctaccagattgctgaagggtgtattaggga 500
                .         .         .         .         .
 651 catactggagcactttccgcacagccatgagaagtcctcttctgtccaaa 700
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 501 catactggagcactttccgcacagccatgagaagtcctcttctgtccaaa 550
                .         .         .         .         .
 701 aacaattacagcctgagcgtttcctggcggatatctattacttccggata 750
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 551 aacaattacagcctgagcgtttcctggcggatatctattacttccggata 600
                .         .         .         .         .
 751 tgcagttacaccgaacaaattgcagtcataaactacatggagaagttcct 800
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 601 tgcagttacaccgaacaaattgcagtcataaactacatggagaagttcct 650
                .         .         .         .         .
 801 cagagagcataaagatgtgcgtatagttattattgatagtgttactttcc 850
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 651 cagagagcataaagatgtgcgtatagttattattgatagtgttactttcc 700
                .         .         .         .         .
 851 actttcgacaagattttgaagatctggcactgaggaccagagtgctaagt 900
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 701 actttcgacaagattttgaagatctggcactgaggaccagagtgctaagt 750
                .         .         .         .         .
 901 ggattatcattgaagttaatgaagattgcaaagacatataacttggcagt 950
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 751 ggattatcattgaagttaatgaagattgcaaagacatataacttggcagt 800
                .         .         .         .         .
 951 tgtcttgttgaaccaagtcactactaaatttacagaagggtcatttcaat 1000
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 801 tgtcttgttgaaccaagtcactactaaatttacagaagggtcatttcaat 850
                .         .         .         .         .
1001 tgactcttgctctaggtgacagctggtcccactcatgcacgaaccggttg 1050
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 851 tgactcttgctctaggtgacagctggtcccactcatgcacgaaccggttg 900
                .         .         .         .         .
1051 attctgcactggaatgggaacgaacgatacgcacatcttgataagtctcc 1100
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 901 attctgcactggaatgggaacgaacgatacgcacatcttgataagtctcc 950
                .         .         .         .         .
1101 ttcacttccagtagcctcagcaccgtatgcagtgacaggcaaagggatta 1150
     |||||||||||||||||||| ||||||||||||||||||||||||||||||
 951 ttcacttccagtagcctcagccccgtatgcagtgacaggcaaagggatta 1000
                .         .         .         .         .
1151 gagatgctgtgagctcaaaccacaagcgagcccgagtaacgtagcattct 1200
     |||||| ||||||||||||||||||||||||||||||||||||||||||||
1001 gagatg.tgtgagctcaaaccacaagcgagcccgagtaacgtagcattct 1049
                .         .         .         .         .
1201 tggtgtcaagcacttgtatgtccactacgctcctgcagctttcttcgcca 1250
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1050 tggtgtcaagcacttgtatgtccactacgctcctgcagctttcttcgcca 1099
                .         .         .         .         .
1251 tggatcttttggactagtgaggtgagactggagaatagtaccattttgtt 1300
```

```
      ||||||||||||||||||||||||||||||||||||||||||||||||      ||
1100 tggatcttttggactagtgaggtgagactggagaatagtaccat....tt 1145
          .         .         .         .         .
1301 gattctcagttgctttgtgccgttggctaccaaccaaccttaagagagaa 1350
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1146 gattctcagttgctttgtgccgttggctaccaaccaaccttaagagagaa 1195
          .         .         .         .         .
1351 gtaaatacaacagaacaggctaatatagtgttttgtatctgaacatctgg 1400
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1196 gtaaatacaacagaacaggctaatatagtgttttgtatctgaacatctgg 1245
          .         .         .         .         .
1401 cccatcgtacattcagtaaagcctataatagcgggca............. 1437
     |||||||||||||||||||||||||||||||||||||
1246 cccatcgtacattcagtaaagcctataatagcgggcatatatgtgcttct 1295
          .         .         .         .         .
1438 ................aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1471
                     ||||||||||||||||||||||||||||||||||
1296 ctgatcaccgatcagcaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1345
          .         .         .         .         .
1472 aaa............................................... 1474
     |||
1346 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1395
                              .
                              .
                              .
```

Input Sequence: 1107sid1

```
!!NA_SEQUENCE 1.0
WPDEF Case 1107 SEQ ID NO: 1
Case 1107 Rad51-like sequences.  From SEQ LISTING.
1107sid1  Length: 1474  August 28, 2001 15:55  Type: N
Check: 4817  ..

       1  tcgacccacg cgtccgcact tgactcccag tctcccactg
tgcgcagttc

      51  gcttggtccc cggagcccca aaggcggcgg tgagccggag
```

View Sequence

Input Sequence: 1107sid3

```
!!NA_SEQUENCE 1.0
WPDEF Case 1107 SEQ ID NO: 3
Case 1107 Rad51-like sequences SEQ ID NO: 3 from SEQ LISTING
1107sid3  Length: 1456  August 28, 2001 15:56  Type: N
Check: 3152  ..

       1  cgacgtaagc ggctgcgtgg cgccaccgac ggaggctacg
agcggttgtg

      51  gaggcagata tgagaggtgg aggtggctac aacgggtcgg
```

View Sequence

Gap Results

GAP of: 1107sid1 check: 4817 from: 1 to: 1474

WPDEF Case 1107 SEQ ID NO: 1
Case 1107 Rad51-like sequences. From SEQ LISTING.

to: 1107sid5 check: 9084 from: 1 to: 1333

WPDEF Case 1107 SEQ ID NO: 5
Case 1107 SEQ ID NO: 5 from SEQ LISTING. Rad51-like sequences

Symbol comparison table: nwsgapdna.cmp CompCheck: 8760

Gap Weight:	50	Average Match:	10.000
Length Weight:	3	Average Mismatch:	0.000
Quality:	13160	Length:	1474
Ratio:	9.872	Gaps:	0
Percent Similarity:	98.725	Percent Identity:	98.650

Match display thresholds for the alignment(s):
| = IDENTITY
: = 5
. = 1

1107sid1 x 1107sid5 August 28, 2001 16:14 ..

```
                                   .
                                   .
                                   .
          .         .         .         .         .
  101 acggcgcggcgcgactccccctaagcgacagcggcggcgtcgacgtaag 150
                                                |||||||||
    1 .........................................cgacgtaag 9
          .         .         .         .         .
  151 cggctgcgtggcgccaccgacggaggctacgagcggttgtggaggcagat 200
      ||||||||||||||||||||||||||||||||||||||||||||||||||
   10 cggctgcgtggcgccaccgacggaggctacgagcggttgtggaggcagat 59
          .         .         .         .         .
  201 atgagaggtggaggtggctacaacgggtcggcggctgtgagatactgaaa 250
      ||||||||||||||||||||||||||||||||||||||||||||||||||
   60 atgagaggtggaggtggctacaacgggtcggcggctgtgagatactgaaa 109
          .         .         .         .         .
  251 tccgcactgcagttctcttcttcccccaatcagtaccacctctccaagtg 300
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  110 tccgcactgcagttctcttcttcccccaatcagtaccacctctccaagtg 159
          .         .         .         .         .
  301 gcaatcaccatgggagatcaatctggctctagaaatggaccacaacagaa 350
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  160 gcaatcaccatgggagatcaatctggctctagaaatggaccacaacagaa 209
          .         .         .         .         .
  351 gtacgtttcaggagcccagaatgcctgggatatgttctctgatgagctgt 400
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  210 gtacgtttcaggagcccagaatgcctgggatatgttctctgatgagctgt 259
          .         .         .         .         .
  401 cacagaaacacatcactactggttctggtgacctcaatgacatacttggt 450
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  260 cacagaaacacatcactactggttctggtgacctcaatgacatacttggt 309
          .         .         .         .         .
  451 ggcgggattcactgcaaagaagttactgagatcggtggcgtcccaggggt 500
      ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
 310 ggcgggattcactgcaaagaagttactgagatcggtggcgtcccaggggt 359

 501 tggtaaaactcaactggggattcaactagcaatcaatgtacaaatcccag 550
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 360 tggtaaaactcaactggggattcaactagcaatcaatgtacaaatcccag 409

 551 tggaatgtggtggccttggtgggaaagcagtttatatagatacagagggc 600
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 410 tggaatgtggtggccttggtgggaaagcagtttatatagatacagagggc 459

 601 agtttcatggttgaacgtgtctaccagattgctgaagggtgtattaggga 650
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 460 agtttcatggttgaacgtgtctaccagattgctgaagggtgtattaggga 509

 651 catactggagcactttccgcacagccatgagaagtcctcttctgtccaaa 700
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 510 catactggagcactttccgcacagccatgagaagtcctcttctgtccaaa 559

 701 aacaattacagcctgagcgtttcctggcggatatctattacttccggata 750
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 560 aacaattacagcctgagcgtttcctggcggatatctattacttccggata 609

 751 tgcagttacaccgaacaaattgcagtcataaactacatggagaagttcct 800
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 610 tgcagttacaccgaacaaattgcagtcataaactacatggagaagttcct 659

 801 cagagagcataaagatgtgcgtatagttattattgatagtgttactttcc 850
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 660 cagagagcataaagatgtgcgtatagttattattgatagtgttactttcc 709

 851 actttcgacaagattttgaagatctggcactgaggaccagagtgctaagt 900
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 710 actttcgacaagattttgaagatctggcactgaggaccagagtgctaagt 759

 901 ggattatcattgaagttaatgaagattgcaaagacatataacttggcagt 950
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 760 ggattatcattgaagttaatgaagattgcaaagacatataacttggcagt 809

 951 tgtcttgttgaaccaagtcactactaaatttacagaagggtcatttcaat 1000
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 810 tgtcttgttgaaccaagtcactactaaatttacagaagggtcatttcaat 859

1001 tgactcttgctctaggtgacagctggtcccactcatgcacgaaccggttg 1050
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 860 tgactcttgctctaggtgacagctggtcccactcatgcacgaaccggttg 909

1051 attctgcactggaatgggaacgaacgatacgcacatcttgataagtctcc 1100
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 910 attctgcactggaatgggaacgaacgatacgcacatcttgataagtctcc 959

1101 ttcacttccagtagcctcagcaccgtatgcagtgacaggcaaagggatta 1150
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 960 ttcacttccagtagcctcagcaccgtatgcagtgacaggcaaagggatta 1009

1151 gagatgctgtgagctcaaaccacaagcgagcccgagtaacgtagcattct 1200
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1010 gagatgctgtgagctcaaaccacaagcgagcccgagtaacgtagcattct 1059

1201 tggtgtcaagcacttgtatgtccactacgctcctgcagctttcttcgcca 1250
     |||||||||||||||||||||||||||||||||| ||||||||||||||||
1060 tggtgtcaagcacttgtatgtccactacgctcctgctgctttcttcgcca 1109

1251 tggatcttttggactagtgaggtgagactggagaatagtaccattttgtt 1300
```

```
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1110 tggatcttttggactagtgaggtgagactggagaatagtaccattttgtt 1159
          .         .         .         .         .
1301 gattctcagttgctttgtgccgttggctaccaaccaaccttaagagagaa 1350
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1160 gattctcagttgctttgtgccgttggctaccaaccaaccttaagagagaa 1209
          .         .         .         .         .
1351 gtaaatacaacagaacaggctaatatagtgttttgtatctgaacatctgg 1400
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1210 gtaaatacaacagaacaggctaatatagtgttttgtatctgaacatctgg 1259
          .         .         .         .         .
1401 cccatcgtacattcagtaaagcctataatagcgggcaaaaaaaaaaaaaa 1450
     :|||||||||||||||||||||||||||||||||||||| | |
1260 sccatcgtacattcagtaaagcctataatagcgggcatatatgtgcttct 1309
          .         .
1451 aaaaaaaaaaaaaaaaaaaaaaa 1474
       |  ||||||||||||||||||
1310 ctgatcaaaaaaaaaaaaaaaaa 1333
```

Input Sequence: 1107sid1

```
!!NA_SEQUENCE 1.0
WPDEF Case 1107 SEQ ID NO: 1
Case 1107 Rad51-like sequences.  From SEQ LISTING.
1107sid1  Length: 1474  August 28, 2001 15:55  Type: N
Check: 4817  ..

       1  tcgacccacg cgtccgcact tgactcccag tctcccactg
tgcgcagttc

      51  gcttggtccc cggagcccca aaggcggcgg tgagccggag
```

View Sequence

Input Sequence: 1107sid5

```
!!NA_SEQUENCE 1.0
WPDEF Case 1107 SEQ ID NO: 5
Case 1107 SEQ ID NO: 5 from SEQ LISTING.  Rad51-like
sequences
1107sid5  Length: 1333  August 28, 2001 16:00  Type: N
Check: 9084  ..

       1  cgacgtaagc ggctgcgtgg cgccaccgac ggaggctacg
agcggttgtg
```

View Sequence

!!NA_MULTIPLE_ALIGNMENT 1.0

Multiple Sequence Alignment Results

Symbol comparison table: pileupdna.cmp CompCheck: 6876

GapWeight: 5
GapLengthWeight: 1

1107sid1_pileup_42431.txt MSF: 1611 Type: N August 28, 2001 16:08 Check: 2750 ..

Name: 1107sid1 Len: 1611 Check: 421 Weight: 1.00
Name: 1107sid5 Len: 1611 Check: 9483 Weight: 1.00
Name: 1107sid3 Len: 1611 Check: 2846 Weight: 1.00

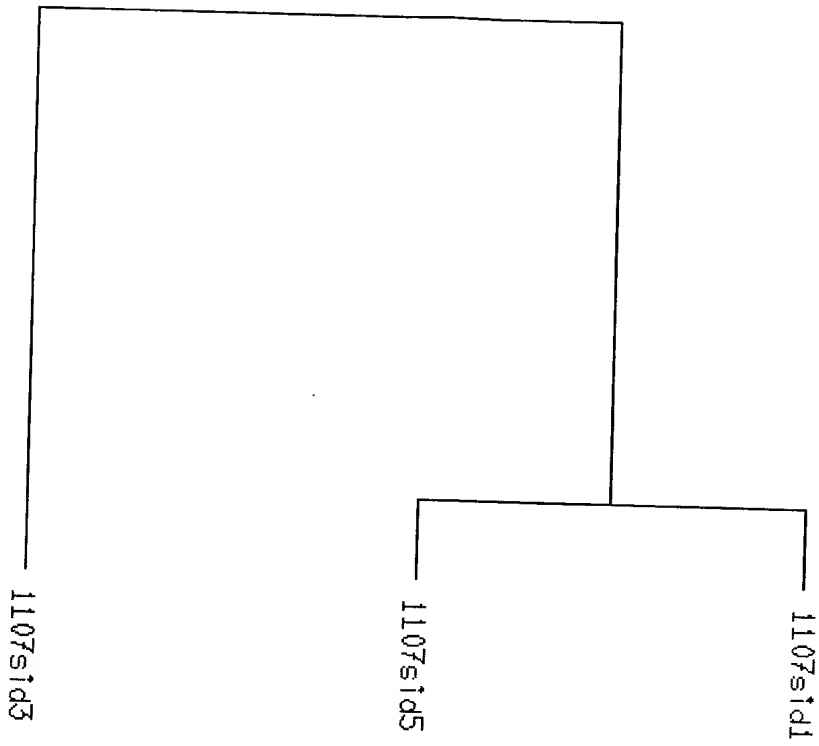
//

```
           1
1107sid1 tcgacccacg cgtccgcact tgactcccag tctcccactg tgcgcagttc
1107sid5 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
1107sid3 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

           51
1107sid1 gcttggtccc cggagcccca aaggcggcgg tgagccggag cccggagacg
1107sid5 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
1107sid3 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

          101
1107sid1 acggcgcggc gcgactcccc cctaagcgac agcggcggcg tcgacgtaag
1107sid5 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~cgacgtaag
1107sid3 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~cgacgtaag

          151
1107sid1 cggctgcgtg gcgccaccga cggaggctac gagcggttgt ggaggcagat
1107sid5 cggctgcgtg gcgccaccga cggaggctac gagcggttgt ggaggcagat
1107sid3 cggctgcgtg gcgccaccga cggaggctac gagcggttgt ggaggcagat

          201
1107sid1 atgagaggtg gaggtggcta caacgggtcg gcggctgtga gatactgaaa
1107sid5 atgagaggtg gaggtggcta caacgggtcg gcggctgtga gatactgaaa
1107sid3 atgagaggtg gaggtggcta caacgggtcg gcggctgtga gatactgaaa

          251
1107sid1 tccgcactgc agttctcttc ttcccccaat cagtaccacc tctccaagtg
1107sid5 tccgcactgc agttctcttc ttcccccaat cagtaccacc tctccaagtg
1107sid3 tccgcactgc agttctcttc ttcccccaat cagtaccacc tctccaagtg

          301
1107sid1 gcaatcacca tgggagatca atctggctct agaaatggac cacaacagaa
1107sid5 gcaatcacca tgggagatca atctggctct agaaatggac cacaacagaa
1107sid3 gcaatcacca tggga...ca atctggctct agaaatggac cacaacagaa

          351
1107sid1 gtacgtttca ggagcccaga atgcctggga tatgttctct gatgagctgt
1107sid5 gtacgtttca ggagcccaga atgcctggga tatgttctct gatgagctgt
1107sid3 gtacgtttca ggagcccaga atgcctggga tatgttctct gatgagctgt

          401
1107sid1 cacagaaaca catcactact ggttctggtg acctcaatga catacttggt
1107sid5 cacagaaaca catcactact ggttctggtg acctcaatga catacttggt
1107sid3 cacagaaaca catcactact ggttctggtg acctcaatga catacttggt
```

```
          451                                                        500
1107sid1 ggcgggattc actgcaaaga agttactgag atcggtggcg tcccaggggt
1107sid5 ggcgggattc actgcaaaga agttactgag atcggtggcg tcccaggggt
1107sid3 ggcgggattc actgcaaaga agttactgag atcggtggcg tcccaggggt

          501                                                        550
1107sid1 tggtaaaact caactgggga ttcaactagc aatcaatgta caaatcccag
1107sid5 tggtaaaact caactgggga ttcaactagc aatcaatgta caaatcccag
1107sid3 tggtaaaact caactgggga ttcaactagc aatcaatgta caaatcccag

          551                                                        600
1107sid1 tggaatgtgg tggccttggt gggaaagcag tttatataga tacagagggc
1107sid5 tggaatgtgg tggccttggt gggaaagcag tttatataga tacagagggc
1107sid3 tggaatgtgg tggccttggt gggaaagcag tttatat... ...agagggc

          601                                                        650
1107sid1 agtttcatgg ttgaacgtgt ctaccagatt gctgaagggt gtattaggga
1107sid5 agtttcatgg ttgaacgtgt ctaccagatt gctgaagggt gtattaggga
1107sid3 agtttcatgg ttgaacgtgt ctaccagatt gctgaagggt gtattaggga

          651                                                        700
1107sid1 catactggag cactttccgc acagccatga gaagtcctct tctgtccaaa
1107sid5 catactggag cactttccgc acagccatga gaagtcctct tctgtccaaa
1107sid3 catactggag cactttccgc acagccatga gaagtcctct tctgtccaaa

          701                                                        750
1107sid1 aacaattaca gcctgagcgt ttcctggcgg atatctatta cttccggata
1107sid5 aacaattaca gcctgagcgt ttcctggcgg atatctatta cttccggata
1107sid3 aacaattaca gcctgagcgt ttcctggcgg atatctatta cttccggata

          751                                                        800
1107sid1 tgcagttaca ccgaacaaat tgcagtcata aactacatgg agaagttcct
1107sid5 tgcagttaca ccgaacaaat tgcagtcata aactacatgg agaagttcct
1107sid3 tgcagttaca ccgaacaaat tgcagtcata aactacatgg agaagttcct

          801                                                        850
1107sid1 cagagagcat aaagatgtgc gtatagttat tattgatagt gttactttcc
1107sid5 cagagagcat aaagatgtgc gtatagttat tattgatagt gttactttcc
1107sid3 cagagagcat aaagatgtgc gtatagttat tattgatagt gttactttcc

          851                                                        900
1107sid1 actttcgaca agattttgaa gatctggcac tgaggaccag agtgctaagt
1107sid5 actttcgaca agattttgaa gatctggcac tgaggaccag agtgctaagt
1107sid3 actttcgaca agattttgaa gatctggcac tgaggaccag agtgctaagt

          901                                                        950
1107sid1 ggattatcat tgaagttaat gaagattgca aagacatata acttggcagt
1107sid5 ggattatcat tgaagttaat gaagattgca aagacatata acttggcagt
1107sid3 ggattatcat tgaagttaat gaagattgca aagacatata acttggcagt

          951                                                       1000
1107sid1 tgtcttgttg aaccaagtca ctactaaatt tacagaaggg tcatttcaat
1107sid5 tgtcttgttg aaccaagtca ctactaaatt tacagaaggg tcatttcaat
1107sid3 tgtcttgttg aaccaagtca ctactaaatt tacagaaggg tcatttcaat

         1001                                                       1050
1107sid1 tgactcttgc tctaggtgac agctggtccc actcatgcac gaaccggttg
1107sid5 tgactcttgc tctaggtgac agctggtccc actcatgcac gaaccggttg
1107sid3 tgactcttgc tctaggtgac agctggtccc actcatgcac gaaccggttg

         1051                                                       1100
1107sid1 attctgcact ggaatgggaa cgaacgatac gcacatcttg ataagtctcc
```

```
1107sid5 attctgcact ggaatgggaa cgaacgatac gcacatcttg ataagtctcc
1107sid3 attctgcact ggaatgggaa cgaacgatac gcacatcttg ataagtctcc

         1101                                                1150
1107sid1 ttcacttcca gtagcctcag caccgtatgc agtgacaggc aaagggatta
1107sid5 ttcacttcca gtagcctcag caccgtatgc agtgacaggc aaagggatta
1107sid3 ttcacttcca gtagcctcag ccccgtatgc agtgacaggc aaagggatta

         1151                                                1200
1107sid1 gagatgctgt gagctcaaac cacaagcgag cccgagtaac gtagcattct
1107sid5 gagatgctgt gagctcaaac cacaagcgag cccgagtaac gtagcattct
1107sid3 gagatg.tgt gagctcaaac cacaagcgag cccgagtaac gtagcattct

         1201                                                1250
1107sid1 tggtgtcaag cacttgtatg tccactacgc tcctgcagct ttcttcgcca
1107sid5 tggtgtcaag cacttgtatg tccactacgc tcctgctgct ttcttcgcca
1107sid3 tggtgtcaag cacttgtatg tccactacgc tcctgcagct ttcttcgcca

         1251                                                1300
1107sid1 tggatctttt ggactagtga ggtgagactg gagaatagta ccattttgtt
1107sid5 tggatctttt ggactagtga ggtgagactg gagaatagta ccattttgtt
1107sid3 tggatctttt ggactagtga ggtgagactg gagaatagta ccat....tt

         1301                                                1350
1107sid1 gattctcagt tgctttgtgc cgttggctac caaccaacct taagagagaa
1107sid5 gattctcagt tgctttgtgc cgttggctac caaccaacct taagagagaa
1107sid3 gattctcagt tgctttgtgc cgttggctac caaccaacct taagagagaa

         1351                                                1400
1107sid1 gtaaatacaa cagaacaggc taatatagtg ttttgtatct gaacatctgg
1107sid5 gtaaatacaa cagaacaggc taatatagtg ttttgtatct gaacatctgg
1107sid3 gtaaatacaa cagaacaggc taatatagtg ttttgtatct gaacatctgg

         1401                                                1450
1107sid1 cccatcgtac attcagtaaa gcctataata gcgggcaaaa aaaaaaaaaa
1107sid5 sccatcgtac attcagtaaa gcctataata gcgggcatat atgtgcttct
1107sid3 cccatcgtac attcagtaaa gcctataata gcgggcatat atgtgcttct

         1451                                                1500
1107sid1 aaaaaaaaaa aaaaaaaaaa aaaa~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
1107sid5 ctgatcaaaa aaaaaaaaaa aaaa~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
1107sid3 ctgatcaccg atcagcaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa

         1501                                                1550
1107sid1 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
1107sid5 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
1107sid3 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa

         1551                                                1600
1107sid1 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
1107sid5 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
1107sid3 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa

         1601       1611
1107sid1 ~~~~~~~~~~ ~
1107sid5 ~~~~~~~~~~ ~
1107sid3 aaaaaaaaaa a
```

Multiple Sequence Alignment Dendrogram August 28, 2001 16:08



Gap Results

GAP of: 1107sid2 check: 4041 from: 1 to: 294

WPDEF Case 1107 SEQ ID NO: 2
Case 1107 SEQ ID NO: 2 from SEQ LISTING. Rad51-like sequences. Protein

to: 1107sid4 check: 3715 from: 1 to: 281

WPDEF Case 1107 SEQ ID NO: 4
Case 1107 SEQ ID NO: 4 from SEQ LISTING. Rad 51-like sequences. Protein

Symbol comparison table: blosum62.cmp CompCheck: 6430
BLOSUM62 amino acid substitution matrix.
Reference: Henikoff, S. and Henikoff, J. G. (1992). Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. USA 89: 10915-10919.

Gap Weight:	8	Average Match:	2.912
Length Weight:	2	Average Mismatch:	-2.003
Quality:	1449	Length:	294
Ratio:	5.157	Gaps:	1
Percent Similarity:	99.644	Percent Identity:	99.644

Match display thresholds for the alignment(s):
| = IDENTITY
: = 2
. = 1

1107sid2 x 1107sid4 August 28, 2001 16:15 ..

```
          .         .         .         .         .
   1 MGDQSGSRNGPQQKYVSGAQNAWDMFSDELSQKHITTGSGDLNDILGGGI 50
     |||||||||||||||||||||||||||||||||||||||||||||||||||
   1 MGDQSGSRNGPQQKYVSGAQNAWDMFSDELSQKHITTGSGDLNDILGGGI 50
          .         .         .         .         .
  51 HCKEVTEIGGVPGVGKTQLGIQLAINVQIPVECGGLGGKAVYIDTEGSFM 100
     |||||||||||||||||||||||||||||||||||||||||||  |||||
  51 HCKEVTEIGGVPGVGKTQLGIQLAINVQIPVECGGLGGKAVYI..EGSFM 98
          .         .         .         .         .
 101 VERVYQIAEGCIRDILEHFPHSHEKSSSVQKQLQPERFLADIYYFRICSY 150
     |||||||||||||||||||||||||||||||||||||||||||||||||||
  99 VERVYQIAEGCIRDILEHFPHSHEKSSSVQKQLQPERFLADIYYFRICSY 148
          .         .         .         .         .
 151 TEQIAVINYMEKFLREHKDVRIVIIDSVTFHFRQDFEDLALRTRVLSGLS 200
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 149 TEQIAVINYMEKFLREHKDVRIVIIDSVTFHFRQDFEDLALRTRVLSGLS 198
          .         .         .         .         .
 201 LKLMKIAKTYNLAVVLLNQVTTKFTEGSFQLTLALGDSWSHSCTNRLILH 250
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 199 LKLMKIAKTYNLAVVLLNQVTTKFTEGSFQLTLALGDSWSHSCTNRLILH 248
          .         .         .         .
 251 WNGNERYAHLDKSPSLPVASAPYAVTGKGIRDAVSSNHKRARVT 294
     |||||||||||||||||||||||||||||||
 249 WNGNERYAHLDKSPSLPVASAPYAVTGKGIRDV..... 281
```

Input Sequence: 1107sid2

```
!!AA_SEQUENCE 1.0
WPDEF Case 1107 SEQ ID NO: 2
Case 1107 SEQ ID NO: 2 from SEQ LISTING.  Rad51-like
sequences. Protein
 1107sid2  Length: 294  August 28, 2001 16:05  Type: P
Check: 4041  ..

       1  MGDQSGSRNG PQQKYVSGAQ NAWDMFSDEL SQKHITTGSG
DLNDILGGGI

      --  ------------------------------------------------
```

View Sequence

Input Sequence: 1107sid4

```
!!AA_SEQUENCE 1.0
WPDEF Case 1107 SEQ ID NO: 4
Case 1107 SEQ ID NO: 4 from SEQ LISTING.  Rad 51-like
sequences. Protein
 1107sid4  Length: 281  August 28, 2001 16:06  Type: P
Check: 3715  ..

       1  MGDQSGSRNG PQQKYVSGAQ NAWDMFSDEL SQKHITTGSG
DLNDILGGGI

      --  ------------------------------------------------
```

View Sequence

Gap Results

GAP of: 1107sid2 check: 4041 from: 1 to: 294

WPDEF Case 1107 SEQ ID NO: 2
Case 1107 SEQ ID NO: 2 from SEQ LISTING. Rad51-like sequences. Protein

to: 1107sid6 check: 4041 from: 1 to: 294

WPDEF Case 1107 SEQ ID NO: 6
Case 1107 SEQ ID NO: 6 from SEQ LISTING. Rad51-like sequences. Protein.

Symbol comparison table: blosum62.cmp CompCheck: 6430
BLOSUM62 amino acid substitution matrix.
Reference: Henikoff, S. and Henikoff, J. G. (1992). Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. USA 89: 10915-10919.

Gap Weight:	8	Average Match:	2.912
Length Weight:	2	Average Mismatch:	-2.003
Quality:	1530	Length:	294
Ratio:	5.204	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

Match display thresholds for the alignment(s):
| = IDENTITY
: = 2
. = 1

1107sid2 x 1107sid6 August 28, 2001 16:16 ..

```
      .         .         .         .         .
    1 MGDQSGSRNGPQQKYVSGAQNAWDMFSDELSQKHITTGSGDLNDILGGGI 50
      ||||||||||||||||||||||||||||||||||||||||||||||||||
    1 MGDQSGSRNGPQQKYVSGAQNAWDMFSDELSQKHITTGSGDLNDILGGGI 50
      .         .         .         .         .
   51 HCKEVTEIGGVPGVGKTQLGIQLAINVQIPVECGGLGGKAVYIDTEGSFM 100
      ||||||||||||||||||||||||||||||||||||||||||||||||||
   51 HCKEVTEIGGVPGVGKTQLGIQLAINVQIPVECGGLGGKAVYIDTEGSFM 100
      .         .         .         .         .
  101 VERVYQIAEGCIRDILEHFPHSHEKSSSVQKQLQPERFLADIYYFRICSY 150
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  101 VERVYQIAEGCIRDILEHFPHSHEKSSSVQKQLQPERFLADIYYFRICSY 150
      .         .         .         .         .
  151 TEQIAVINYMEKFLREHKDVRIVIIDSVTFHFRQDFEDLALRTRVLSGLS 200
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  151 TEQIAVINYMEKFLREHKDVRIVIIDSVTFHFRQDFEDLALRTRVLSGLS 200
      .         .         .         .         .
  201 LKLMKIAKTYNLAVVLLNQVTTKFTEGSFQLTLALGDSWSHSCTNRLILH 250
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  201 LKLMKIAKTYNLAVVLLNQVTTKFTEGSFQLTLALGDSWSHSCTNRLILH 250
      .         .         .         .
  251 WNGNERYAHLDKSPSLPVASAPYAVTGKGIRDAVSSNHKRARVT 294
      |||||||||||||||||||||||||||||||||||||||||||
  251 WNGNERYAHLDKSPSLPVASAPYAVTGKGIRDAVSSNHKRARVT 294
```

Input Sequence: 1107sid2

```
!!AA_SEQUENCE 1.0
WPDEF Case 1107 SEQ ID NO: 2
Case 1107 SEQ ID NO: 2 from SEQ LISTING.  Rad51-like
sequences. Protein
 1107sid2  Length: 294  August 28, 2001 16:05  Type: P
Check: 4041  ..

       1  MGDQSGSRNG PQQKYVSGAQ NAWDMFSDEL SQKHITTGSG
DLNDILGGGI
```

View Sequence

Input Sequence: 1107sid6

```
!!AA_SEQUENCE 1.0
WPDEF Case 1107 SEQ ID NO: 6
Case 1107 SEQ ID NO: 6 from SEQ LISTING. Rad51-like
sequences. Protein.
 1107sid6  Length: 294  August 28, 2001 16:07  Type: P
Check: 4041  ..

       1  MGDQSGSRNG PQQKYVSGAQ NAWDMFSDEL SQKHITTGSG
DLNDILGGGI
```

View Sequence

!!AA_MULTIPLE_ALIGNMENT 1.0

Multiple Sequence Alignment Results

Symbol comparison table: blosum62.cmp CompCheck: 6430

GapWeight: 8
GapLengthWeight: 2

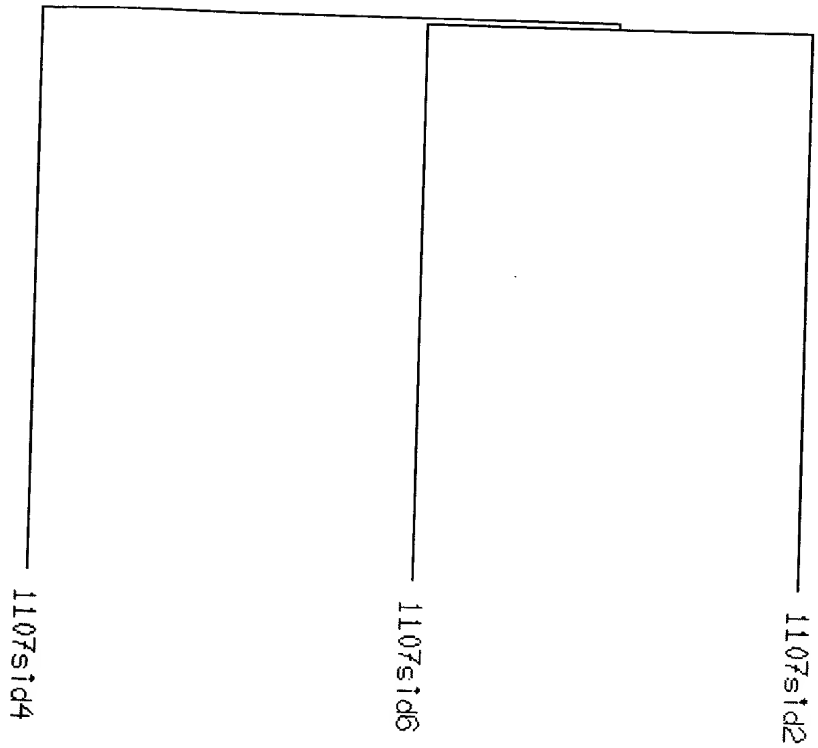
1107sid2_pileup_42538.txt MSF: 294 Type: P August 28, 2001 16:10 Check: 7808 ..

Name: 1107sid2	Len:	294	Check: 4041	Weight:	1.00
Name: 1107sid6	Len:	294	Check: 4041	Weight:	1.00
Name: 1107sid4	Len:	294	Check: 9726	Weight:	1.00

//

```
            1
1107sid2 MGDQSGSRNG PQQKYVSGAQ NAWDMFSDEL SQKHITTGSG DLNDILGGGI
1107sid6 MGDQSGSRNG PQQKYVSGAQ NAWDMFSDEL SQKHITTGSG DLNDILGGGI
1107sid4 MGDQSGSRNG PQQKYVSGAQ NAWDMFSDEL SQKHITTGSG DLNDILGGGI

            51
1107sid2 HCKEVTEIGG VPGVGKTQLG IQLAINVQIP VECGGLGGKA VYIDTEGSFM
1107sid6 HCKEVTEIGG VPGVGKTQLG IQLAINVQIP VECGGLGGKA VYIDTEGSFM
1107sid4 HCKEVTEIGG VPGVGKTQLG IQLAINVQIP VECGGLGGKA VYI..EGSFM

           101
1107sid2 VERVYQIAEG CIRDILEHFP HSHEKSSSVQ KQLQPERFLA DIYYFRICSY
1107sid6 VERVYQIAEG CIRDILEHFP HSHEKSSSVQ KQLQPERFLA DIYYFRICSY
1107sid4 VERVYQIAEG CIRDILEHFP HSHEKSSSVQ KQLQPERFLA DIYYFRICSY

           151
1107sid2 TEQIAVINYM EKFLREHKDV RIVIIDSVTF HFRQDFEDLA LRTRVLSGLS
1107sid6 TEQIAVINYM EKFLREHKDV RIVIIDSVTF HFRQDFEDLA LRTRVLSGLS
1107sid4 TEQIAVINYM EKFLREHKDV RIVIIDSVTF HFRQDFEDLA LRTRVLSGLS

           201
1107sid2 LKLMKIAKTY NLAVVLLNQV TTKFTEGSFQ LTLALGDSWS HSCTNRLILH
1107sid6 LKLMKIAKTY NLAVVLLNQV TTKFTEGSFQ LTLALGDSWS HSCTNRLILH
1107sid4 LKLMKIAKTY NLAVVLLNQV TTKFTEGSFQ LTLALGDSWS HSCTNRLILH

           251
1107sid2 WNGNERYAHL DKSPSLPVAS APYAVTGKGI RDAVSSNHKR ARVT
1107sid6 WNGNERYAHL DKSPSLPVAS APYAVTGKGI RDAVSSNHKR ARVT
1107sid4 WNGNERYAHL DKSPSLPVAS APYAVTGKGI RDV~~~~~~~ ~~~~
```

Multiple Sequence Alignment Dendrogram August 28, 2001 16:1(



APPENDIX B

Gap Results

GAP of: 1107sid1 check: 4817 from: 1 to: 1474

WPDEF Case 1107 SEQ ID NO: 1
Case 1107 Rad51-like sequences. From SEQ LISTING.

to: ac002387cds check: 3310 from: 1 to: 999

WPDEF Case 1107 At Rad51
AC002387 chromosome 2.
Locus AAB82635
GI 2583126

Symbol comparison table: nwsgapdna.cmp CompCheck: 8760

Gap Weight:	50	Average Match:	10.000
Length Weight:	3	Average Mismatch:	0.000
Quality:	5743	Length:	1475
Ratio:	5.749	Gaps:	4
Percent Similarity:	60.120	Percent Identity:	60.120

Match display thresholds for the alignment(s):
| = IDENTITY
: = 5
. = 1

1107sid1 x ac002387cds August 30, 2001 11:35 ..

```
151 cggctgcgtggcgccaccgacggaggctacgagcggttgtggaggcagat 200
                            | ||     | |||  |  |
  1 ..........................atgatttcatttgggcggcgta 22

201 atgagaggtggaggtggctacaacgggtcggcggctgtgagatactgaaa 250
    |   | |  ||       |       |||     |  |  ||   |  |
 23 aatcgccggcgattgaagaaacttcactcgcgacttcagtcatggaggca 72

251 tccgcactgcagttctcttcttcccccaatcagtaccacctctccaagtg 300
    |      | | ||| ||  ||||    |                   ||
 73 tggaggttaccgttatcgccttcgatta.....gaggaaaact 110

301 gcaatcaccatgggagatcaatctggc.tctagaaatggaccacaacaga 349
    |  |||  |  |  | |    |||| | || |    |
111 gatatcggccggttatacttgtctgtcttcgattgcttccgtctcttctt 160

350 agtacgtttcaggagcccagaatgcctgggatatgttctctgatgagctg 399
      |  |  |  ||||  |||| || |||||||||  |    || |||  |
161 ctgatctcgctcgagcaaagaacgcttgggatatgcttcacgaggaggag 210

400 tcacagaaacacatcactactggttctggtgacctcaatgacatacttgg 449
    ||   |   |  || |||||   ||    ||| ||  || ||||  | ||
211 tctttgccgcgtattactacatcttgctctgatcttgataacattttggg 260

450 tggcgggattcactgcaaagaagttactgagatcggtggcgtcccagggg 499
     || || |||  ||| |  || ||||| ||||| ||||| || ||||||
261 cggtggaattagctgtagggatgttacagagattggtggggtaccaggga 310
```

```
 500 ttggtaaaactcaactgggattcaactagcaatcaatgtacaaatccca 549
     |||| || |||||  | |||||| || ||  |  | ||||| || || |||
 311 ttggcaagactcagattgggatccagctctctgtgaatgttcagattcca 360
           .         .         .         .         .
 550 gtggaatgtggtggccttggtgggaaagcagtttatatagatacagaggg 599
      || |||||||| ||||| |||||||| | ||||| |||||||| ||
 361 cgtgagtgtggtggtcttggagggaaagctatatatatcgatacagaagg 410
           .         .         .         .         .
 600 cagtttcatggttgaacgtgtctaccagattgctgaagggtgtattaggg 649
      || |||||||| || ||||  |  ||||| || ||||  ||| |   ||
 411 tagcttcatggtggagcgtgctttacagatagcagaagcttgtgtagagg 460
           .         .         .         .         .
 650 acatactggagcactttccgcacagccatgagaagtcctcttctgtccaa 699
     ||||    ||  ||       |||  ||| |  | |      |    |||
 461 acatggaagaatacacaggatacatgcataaacattttcaagcaaatcaa 510
           .         .         .         .         .
 700 aaacaattacagcctgagcgtttcctggcggatatctattacttccggat 749
       |||| |  | || ||   | || | | | | || |  ||||||||  |
 511 gtacaaatgaaaccagaagatatcttagagaacatattctacttccgtgt 560
           .         .         .         .         .
 750 atgcagttacaccgaacaaattgcagtcataaactacatggagaagttcc 799
      |||||||||||||| ||||| ||| |  | ||  |  | || ||||||
 561 ctgcagttacaccgagcaaatcgcattggtcaatcatcttgaaaagttca 610
           .         .         .         .         .
 800 tcagagagcataaagatgtgcgtatagttattattgatagtgttactttc 849
     ||   ||  | ||||||||     | || ||  | || ||| | || ||
 611 tctctgaaaacaaagatgt...agttgtaatcgtagacagtatcaccttt 657
           .         .         .         .         .
 850 cactttcgacaagattttgaagatctggcactgaggaccagagtgctaag 899
     || || || || || | ||| ||  | || | ||||||  ||||||| ||
 658 catttccgtcaggactatgatgacttagcccagaggacacgagtgctcag 707
           .         .         .         .         .
 900 tggattatcattgaagttaatgaagattgcaaagacatataacttggcag 949
      | | |  | || |||||| |||||| |||| || |  |      | || |
 708 cgaaatggctttaaagttcatgaagcttgccaaaaagttctcacttgcgg 757
           .         .         .         .         .
 950 ttgtcttgttgaaccaagtcactactaaatttacagaagggtcatttcaa 999
     | || ||  | ||||| || || || || ||||  ||||| || ||||||
 758 tcgtgttactaaaccaggtgaccacaaagtttagtgaaggctcgtttcaa 807
           .         .         .         .         .
1000 ttgactcttgctctaggtgacagctggtcccactcatgcacgaaccggtt 1049
      |  | |||||| |||| || |||||||| || || ||||| |||||  |
 808 ctagcgcttgctttaggcgatagctggtctcattcgtgcaccaaccgagt 857
           .         .         .         .         .
1050 gattctgcactggaatgggaacgaacgatacgcacatcttgataagtctc 1099
      |||||| | |||||||||  | || || |||||| || | |||||||| |
 858 cattctgtattggaatggtgatgagcgttacgcatatatcgataagtccc 907
           .         .         .         .         .
1100 cttcacttccagtagcctcagcaccgtatgcagtgacaggcaaagggatt 1149
     ||||||||||   ||| || ||  | ||  | || ||  | | |||  |
 908 cttcacttccttcagcttcggcttcatacactgtaaccagtagaggtcta 957
           .         .         .         .         .
1150 agagatgctgtgagctcaaaccacaagcgagcccgagtaacgtagcattc 1199
     ||| |  |      ||| |    |||||||| |    | | |||
 958 agaaactc...atcctcgagtagcaagcgagtcaagatgatgtaa..... 999
                         .
                         .
                         .
```

Input Sequence: 1107sid1

```
!!NA_SEQUENCE 1.0
WPDEF Case 1107 SEQ ID NO: 1
Case 1107 Rad51-like sequences.  From SEQ LISTING.
1107sid1  Length: 1474  August 28, 2001 15:55  Type: N
Check: 4817  ..

       1  tcgacccacg cgtccgcact tgactcccag tctcccactg
tgcgcagttc

      51  gcttggtccc cggagcccca aaggcggcgg tgagccggag
```

View Sequence

Input Sequence: ac002387cds

```
!!NA_SEQUENCE 1.0
WPDEF Case 1107 At Rad51
AC002387 chromosome 2.
Locus AAB82635
GI 2583126
ac002387cds  Length: 999  August 30, 2001 11:30  Type: N
Check: 3310  ..

       1  atgatttcat ttgggcggcg taaatcgccg gcgattgaag
aaacttcact
```

View Sequence

FrameAlign Results

Local alignment of: 1107sid1 check: 4817 from: 1 to: 1474

WPDEF Case 1107 SEQ ID NO: 1
Case 1107 Rad51-like sequences. From SEQ LISTING.

to: 1107ac002387pep check: 9453 from: 1 to: 332

WPDEF Case 1107 At Rad51 protein
AC002387 protein
Locus AAB82635
GI 2583126
Case 1107 Rad51

Scoring matrix: blosum62.cmp
CompCheck: 1102
BLOSUM62 amino acid substitution matrix.
Reference: Henikoff, S. and Henikoff, J. G. (1992). Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. USA 89: 10915-10919.
Translation table: transl_table_01.txt
transl_table = 1
This file contains the Standard Code specified in the feature table definition, Version 1.08, formatted for use with the GCG programs (Data Files volume of the Data Reference Set). It names amino acids in both one and three-letter form and lists the codons which should translate into them. All GCG translation programs may generate their . . .

Gap Weight:	8	Average Match:	2.778
Length Weight:	2	Average Mismatch:	-2.248
Frameshift Weight:	0		
Quality:	982	Length:	813
Ratio:	3.637	Gaps:	1
Percent Similarity:	78.519	Percent Identity:	67.037

Match display thresholds for the alignment(s):
| = IDENTITY
: = 2
. = 1

1107sid1 x 1107ac002387pep February 25, 2002 17:01 ..

```
                .         .         .         .         .
    364 gcccagaatgcctgggatatgttctctgatgagctgtcacagaaacacat 413
        |||...|||||||||||||||      :::|||   |||         ||
     59 AlaLysAsnAlaTrpAspMetLeuHisGluGluGluSerLeuProArgIl 75
                .         .         .         .         .
    414 cactactggttctggtgacctcaatgacatacttggtggcgggattcact 463
        |||||||         ||||||......||||||||||||||||   |
     76 eThrThrSerCysSerAspLeuAspAsnIleLeuGlyGlyGlyIleSerC 92
                .         .         .         .         .
    464 gcaaagaagttactgagatcggtggcgtcccaggggttggtaaaactcaa 513
        ||::::::|||||||||||||||||||||||||:::|||||||||||||
     93 ysArgAspValThrGluIleGlyGlyValProGlyIleGlyLysThrGln 108
                .         .         .         .         .
    514 ctggggattcaactagcaatcaatgtacaaatcccagtggaatgtggtgg 563
        :::|||||||||||||...:::|||||||||||||||   ||||||||||
    109 IleGlyIleGlnLeuSerValAsnValGlnIleProArgGluCysGlyGl 125
                .         .         .         .         .
```

```
 564 ccttggtgggaaagcagtttatatagatacagagggcagtttcatggttg 613
     ||||||||||||||||:::|||||||||||||||||||||||||||||||
 126 yLeuGlyGlyLysAlaIleTyrIleAspThrGluGlySerPheMetValG 142

 614 aacgtgtctaccagattgctgaagggtgtattagggacatactggagcac 663
     |||||      ||||||||||||   |||:::   |||...   |||:::
 143 luArgAlaLeuGlnIleAlaGluAlaCysValGluAspMetGluGluTyr 158

 664 tttccgcacagccatgagaagtcctcttctgtccaaaaacaattacagcc 713
         :::   |||...          ...   |||   |||::::...||
 159 ThrGlyTyrMetHisLysHisPheGlnAlaAsnGlnValGlnMetLysPr 175

 714 tgagcgtttcctggcggatatctattacttccggatatgcagttacaccg 763
     ||||      |||   ...|||:::|||||||||:::|||||||||||||
 176 oGluAspIleLeuGluAsnIlePheTyrPheArgValCysSerTyrThrG 192

 764 aacaaattgcagtcataaactacatggagaagttcctcagagagcataaa 813
     ||||||||||||...:::::|||::::::|||||||||:::   |||...|||
 193 luGlnIleAlaLeuValAsnHisLeuGluLysPheIleSerGluAsnLys 208

 814 gatgtgcgtatagttattattgatagtgttactttccactttcgacaaga 863
     ||||||   :::||||||:::||||||:::||||||||||||||||||||
 209 AspVal...ValValIleValAspSerIleThrPheHisPheArgGlnAs 224

 864 ttttgaagatctggcactgaggaccagagtgctaagtggattatcattga 913
     |::::::|||||||||   ||||||||||||||||||   :::...||||
 225 pTyrAspAspLeuAlaGlnArgThrArgValLeuSerGluMetAlaLeuL 241

 914 agttaatgaagattgcaaagacatataacttggcagttgtcttgttgaac 963
     ||   ||||||:::||||||   ::::...||||||||||||||||||||
 242 ysPheMetLysLeuAlaLysLysPheSerLeuAlaValValLeuLeuAsn 257

 964 caagtcactactaaatttacagaagggtcatttcaattgactcttgctct 1013
     ||||||||||||||||||...||||||||||||||||   ||||||||
 258 GlnValThrThrLysPheSerGluGlySerPheGlnLeuAlaLeuAlaLe 274

1014 aggtgacagctggtcccactcatgcacgaaccggttgattctgcactgga 1063
     |||||||||||||||||||||||||||||||...||||||:::||||
 275 uGlyAspSerTrpSerHisSerCysThrAsnArgValIleLeuTyrTrpA 291

1064 atgggaacgaacgatacgcacatcttgataagtctccttcacttccagta 1113
     |||||...||||||||||||:::::||||||||||||||||||||||
 292 snGlyAspGluArgTyrAlaTyrIleAspLysSerProSerLeuProSer 307

1114 gcctcagcaccgtatgcagtgacaggcaaagggattagagatgctgtgag 1163
     |||||||||   |||   ||||||   :::|||:::|||.....   ||
 308 AlaSerAlaSerTyrThrValThrSerArgGlyLeuArgAsnSerSerSe 324

1164 ctcaaaccacaag 1176
     ||||...   :::
 325 rSerSerLysArg 328
```

Input Sequence: 1107sid1

```
!!NA_SEQUENCE 1.0
WPDEF Case 1107 SEQ ID NO: 1
Case 1107 Rad51-like sequences.  From SEQ LISTING.
1107sid1  Length: 1474  August 28, 2001 15:55  Type: N
Check: 4817  ..

       1  tcgacccacg cgtccgcact tgactcccag tctcccactg
tgcgcagttc

      51  gcttggtccc cggagcccca aaggcggcgg tgagccggag
```

View Sequence

Input Sequence: 1107ac002387pep

```
!!AA_SEQUENCE 1.0
WPDEF Case 1107 At Rad51 protein
AC002387 protein
Locus AAB82635
GI 2583126
Case 1107 Rad51
ac002387pep  Length: 332  August 30, 2001 11:32  Type: P
Check: 9453  ..

       1  MISFGRRKSP AIEETSLATS VMEAWRLPLS PSIRGKLISA
```

View Sequence

APPENDIX C

PDB
PROTEIN DATA BANK

Query Result Browser

Your query found **45** structures in the current PDB release and you have selected **6** structures so far. Only the selected structures are currently shown. To examine an individual structure select the Explore link!

Pull down to select option: New Search Go

☑ **1EW1** *Deposited:* **21-Apr-2000** *Exp. Method:* **NMR, 10 Structures** { **EXPLORE** }
Title Reca Protein-Bound Single-Stranded DNA
Classification Deoxyribonucleic Acid
Compound Mol_Id: 1; Molecule: DNA (5'-D(TpApCpG)-3'); Chain: A; Engineered: Yes; Other_Details: Reca Protein-Bound Single-Stranded DNA

☑ **1G18** *Deposited:* **11-Oct-2000** *Exp. Method:* **X-ray Diffraction** *Resolution:* **3.80 Å** { **EXPLORE** }
Title Reca-ADP-Alf4 Complex
Classification Hydrolase
Compound Mol_Id: 1; Molecule: Reca Protein; Chain: A; Synonym: Recombination Protein Reca; Ec: 3.4.99.37

☑ **1G19** *Deposited:* **11-Oct-2000** *Exp. Method:* **X-ray Diffraction** *Resolution:* **3.00 Å** { **EXPLORE** }
Title Structure Of Reca Protein
Classification Hydrolase
Compound Mol_Id: 1; Molecule: Reca Protein; Chain: A; Synonym: Recombination Protein Reca; Ec: 3.4.99.37

☑ **1REA** *Deposited:* **19-Dec-1991** *Exp. Method:* **X-ray Diffraction** *Resolution:* **2.70 Å** { **EXPLORE** }
Title Structure Of The Reca Protein-ADP Complex
Classification DNA Binding Protein
Compound Reca Protein (E.C. 3.4.99.37) Complex With Adenosine Diphosphate (Reca-ADP)

☑ **2REB** *Deposited:* **06-Mar-1992** *Exp. Method:* **X-ray Diffraction** *Resolution:* **2.30 Å** { **EXPLORE** }
Title The Structure Of The E. Coli Reca Protein Monomer and Polymer
Classification DNA Binding Protein
Compound Reca Protein (E.C. 3.4.99.37)

Proc. Natl. Acad. Sci. USA
Vol. 94, pp. 6623–6628, June 1997
Biochemistry

An extended DNA structure through deoxyribose-base stacking induced by RecA protein

(homologous genetic recombination/NMR/NMR spectroscopy/transferred nuclear Overhauser effect)

TARO NISHINAKA*†, YUTAKA ITO*, SHIGEYUKI YOKOYAMA†‡, AND TAKEHIKO SHIBATA*§

*Cellular and Molecular Biology Laboratory, and ‡Cellular Signaling Laboratory, The Institute of Physical and Chemical Research (RIKEN), Saitama 351-01, Japan; and †Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, Tokyo 113, Japan

Communicated by Charles R. Cantor, Boston University, Boston, MA, April 14, 1997 (received for review December 2, 1996)

ABSTRACT **The family of proteins that are homologous to RecA protein of *Escherichia coli* is essential to homologous genetic recombination in various organisms including viruses, bacteria, lower eukaryotes, and mammals. In the presence of ATP (or ATPγS), these proteins form helical filaments containing single-stranded DNA at the center. The single-stranded DNA bound to RecA protein is extended 1.5 times relative to B-form DNA with the same sequence, and the extension is critical to pairing with homologous double-stranded DNA. This pairing reaction, called homologous pairing, is a key reaction in homologous recombination. In this NMR study, we determined a three-dimensional structure of the single-stranded DNA bound to RecA protein. The DNA structure contains novel deoxyribose-base stacking in which the 2′-methylene moiety of each deoxyribose is placed above the base of the following residue, instead of normal stacking of adjacent bases. As a result of this deoxyribose-base stacking, bases of the single-stranded DNA are spaced out nearly 5 Å. Thus, this novel structure well explains the axial extension of DNA in the RecA-filaments relative to B-form DNA and leads to a possible interpretation of the role of this extension in homologous pairing.**

Homologous genetic recombination plays critical roles in both evolution and maintenance of a functional genome. RecA protein is essential to homologous recombination in *Escherichia coli* (1, 2), and promotes ATP-dependent joint-molecule formation from homologous double-stranded DNA and single-stranded DNA through "homologous pairing" *in vitro* (3, 4). Homologous pairing by RecA protein has been extensively studied for more than a decade. How single-stranded DNA recognizes sequence homology in double-stranded DNA has been a central question in these studies. Based on studies using chemical probing, electron microscopy, modification of base sequences, mutant RecA proteins, and others, various models such as triplex formation have been proposed to explain the mechanism of recognition of homology (see refs. 5–12 for reviews). However, little information is available on the three-dimensional structures of DNA during homologous pairing, information that is essential for a clear view of the mechanism of homologous recognition.

At the first stage of homologous pairing, RecA protein binds to single-stranded DNA in the presence of ATP, and then double-stranded DNA binds to the nucleoprotein complex for searching for homology (13, 14). Electron microscopic studies revealed that RecA protein forms helical filaments on the single-stranded DNA. Biochemical studies showed that such filaments formed in the presence of ATP ("presynaptic filament") are molecular machines for homologous pairing of the single-stranded DNA in the filaments with naked double-stranded DNA that is then taken up into the filament (14, 15). Under certain conditions, RecA protein forms a filament on double-stranded DNA, whose shape is very similar to that of the filament formed on single-stranded DNA (14, 16, 17). In these RecA filaments, both single-stranded and double-stranded DNA are extended 1.5 times as compared with B-form DNA. In spite of low degrees of amino acid sequence homology, eukaryotic homologs of RecA protein, the Rad51 proteins from *Saccharomyces cerevisiae* and *Homo sapiens*, and the functional homolog UvsX protein from *coli-phage* T4 form helical nucleoprotein filaments that have a shape that is nearly identical to bacterial RecA protein, as revealed by electron microscopy (18–20). In the experiments described here, we determined a three-dimensional structure of single-stranded DNA bound to RecA protein, which revealed a novel stacking of deoxyribose and bases.

MATERIALS AND METHODS

Oligodeoxyribonucleotides and RecA Protein. Oligonucleotides were synthesized on a DNA synthesizer (EXPEDITE; Millipore) followed by the purification with reversed-phase column cartridges (Oligo-pak SP; Millipore), or purchased from Cruachem (Kyoto) or Genset (Tokyo). Undesirable organic impurities and metal ions were removed by using cation ion exchange resins (AG 50W-X8, Chelex 100; Bio-Rad). The purified oligonucleotides were lyophilized rapidly and stored at −20°C. DNA concentrations were determined by absorbance measurements at 260 nm and are expressed in moles of entire molecules rather than moles of nucleotide residues.

RecA protein was purified as described by Shibata *et al.* (21, 22), with a minor modification, and dialyzed against 20 mM Tris·Cl (pH 7.5) buffer containing 6.7 mM $MgCl_2$ and 150 mM NaCl. By use of ultrafiltration, we concentrated RecA protein and replaced the solvent by a deuterium buffer {20 mM [uniform ^2H] Tris·Cl, pH 7.1 (pH values were uncorrected for isotope effects)/6.7 mM $MgCl_2$/150 mM NaCl}; i.e., the protein solution was centrifuged at 3,000 rpm for 0.5–2 hr at 4°C in a Centriprep cartridge (30-kDa cut-off; Amicon), followed by dilution with the deuterium buffer. We repeated this process several times. The sample was then lyophilized, stored at −20°C, and redissolved in D_2O (99.96%; Euriso-top) before use. The activity of the preparation of RecA protein was

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; TRNOE, transferred NOE.

Data deposition: The atomic coordinates have been deposited in the Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Upton, NY 11973 (3rec).

§To whom reprint requests should be addressed at: Cellular and Molecular Biology Laboratory, The Institute of Physical and Chemical Research (RIKEN), Wako-shi, Saitama 351-01, Japan. e-mail: tshibata@postman.riken.go.jp.

assessed by assaying the single-stranded DNA-dependent ATPase activity, which was not changed by lyophilization.

Just after the purification, the concentrations of RecA protein were first determined by the Folin phenol–reagent method described by Lowry *et al.* (23), with bovine serum albumin as a standard. After preparation for NMR spectroscopic observations, as just described, the concentrations were determined again by the Bradford method (Bio-Rad), with untreated RecA protein as the standard. The amounts of RecA protein are expressed as moles of 38-kDa polypeptide.

NMR Spectroscopy. One- and two-dimensional spectra were measured on a Bruker AMX600 spectrometer at 20°C–37°C in 20 mM [uniform ^{2}H] Tris·Cl buffer (pH 7.0) containing 6.7 mM $MgCl_2$ and 150 mM NaCl in D_2O.

For one-dimensional nuclear Overhauser effect (NOE) difference spectra an objective proton was irradiated for 0.5 sec prior to a 90° read pulse, and the free induction decays were subtracted from those of off-irradiated scans. The water signal was presaturated for 1.5 sec. The spectra were recorded with 6,024 Hz of spectra width and 16 *k* data points. Total measuring time was 29 min.

To attenuate undesirable protein resonances in the transferred NOE spectroscopy (NOESY) spectrum, we applied a short $T_{1\rho}$ filter (20 ms, $\gamma B_1 = 3$ kHz) before a standard NOESY pulse scheme (pre-sat. $-90°_x - SL_y - 90°_{-x} - g - 90° - t_1 - 90° - \tau - g - 90° -$ Acq.; g, a gradient pulse; ref. 24). The mixing time was varied randomly over 8% of the designated mixing time to suppress zero-quantum artifacts in transferred NOESY spectra (25). The spectra were acquired with a total of 1,024 (t_2) × 400 (t_1) complex points and spectral width of 6,024 Hz. The free induction decays were apodized with a 90°-shifted skewed sinebell function (skew parameter 2) before Fourier transformation in both dimensions. All two-dimensional data sets were processed and analyzed with the program FELIX, version 2.3 (Biosym Technologies, San Diego) using Silicon Graphics workstations.

Analysis of NMR Data and Structural Calculation. NOE intensities in two-dimensional spectra at mixing times of 100, 120, 180, or 200 msec were integrated and calibrated using those of intra-residue C[H5]–C[H6] crosspeaks as references, except in the case of d(TAG). For d(TAG), the intensity of H2′–H1′ crosspeak was used as a reference. Interproton distance of H2′–H1′ holds a nearly constant value between C2′-*endo* (2.99 Å) and C3′-*endo* (2.73 Å). All intensities were converted to distance restraints using the $I/I_0 = (r_{ij}/r_0)^{-6}$ relation, where I = peak intensity, r_{ij} = distance between i and j protons, I_0 = peak intensity of the reference and r_0 = distance of the reference. All restraints were classified as short ($r \leq 3.5$ Å), medium (3.5 Å $< r \leq 4.5$ Å), and long (4.5 Å $< r$) and set the upper and lower bound ± 0.3 Å, ±0.4 Å, and ± 0.5 Å, respectively. The additional distance restraints appeared in longer mixing time (≤ 200 msec) and repulsive distance restraints were incorporated during the structural refinement.

All structural calculations were carried out by the use of X-PLOR, version 3.1 (26). The refinement protocol followed essentially the procedures outlined in the X-PLOR manual. After a short cycle of energy minimization, simulated annealing calculations were initiated at 1,000 K and run for 18 ps. The temperature was then lowered to 100 K with 25 K step size and 1.5 ps dynamics during each cooling step. The structure was then minimized for 200 cycles. Because the sugar puckering was ill determined, the obtained structure was further minimized by use of the revised parameters published by Parkinson *et al.* (27).

Assignments for all nonexchangeable protons of DNA (except stereospecific assignments of H5′ and H5″) were obtained by analysis of double quantum filtered correlated spectroscopy (DQF-COSY), total correlated spectroscopy (TOCSY), rotating-frame Overhauser effect spectroscopy (ROESY), and ^{1}H-^{31}P correlated spectroscopy (^{1}H-^{31}P COSY) spectra using standard sequential assignment techniques (28). Stereospecific assignments of H5′ and H5″ were carried out during the process of structural refinement.

RESULTS

Transferred NOE Analysis of Single-Stranded Oligodeoxyribonucleotides Bound to RecA Protein. The structure of single-stranded DNA induced by the binding to RecA protein in the presence of ATPγS was analyzed by means of the transferred NOE (TRNOE) using short (3–6-mer) oligodeoxyribonucleotides: d(TAG), d(CGA), d(TACG), and d(TGACAT). The TRNOE allows us to analyze structures of small ligands bound to large molecules when the exchange between bound and free states is fast enough (29–33).

We added RecA protein stepwise to the solution containing oligodeoxyribonucleotides and ATPγS (an unhydrolyzable ATP analog). The chemical shifts of the resonances were slightly moved and the signals were slightly broadened in one-dimensional ^{1}H-NMR spectra after the addition of RecA protein. We observed no resonances derived from the bound state, which should have appeared if the exchange rate were slow. Signals from RecA protein were hardly detected due to

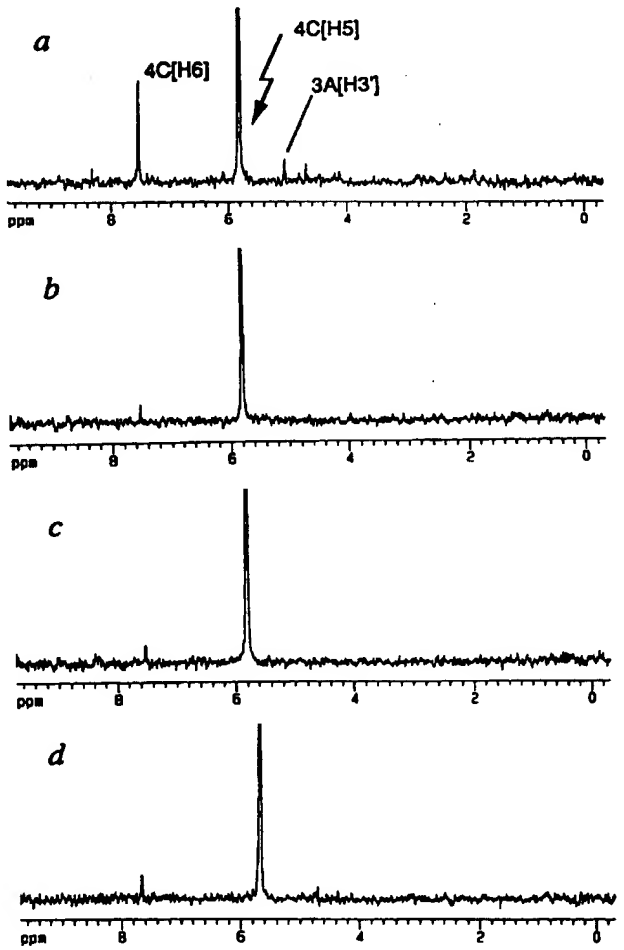


FIG. 1. One-dimensional TRNOE difference spectra of an oligodeoxyribonucleotide. (*a* and *b*) Spectra of 1.1 mM d(TGACAT) and 54 μM RecA protein in the presence of 1.1 mM ATPγS (*a*) or ADP (*b*) in D_2O at 37°C. (*c*) Spectrum of the DNA solution before the addition of RecA protein and ATPγS or ADP. (*d*) DNA was replaced by RNA; 1.1 mM r(UGACAU) in the presence of RecA protein and ATPγS. The cytosine H5 proton of DNA or RNA was irradiated for 0.5 sec before a 90° read pulse.

severe signal broadening. From $T_{1\rho}$ measurements as a function of the spin-lock field strength, we have determined the dissociation rate constants for the oligodeoxyribonucleotide–RecA complex. The value for d(TGACAT) at 30°C was 40,000 (±4,000) s^{-1}, which would be fast enough compared with the chemical shift scale and the cross-relaxation rate (T.N. and Y.I., unpublished observation).

ATP is an essential cofactor for RecA protein-mediated homologous pairing. ATP is hydrolyzed by RecA protein during the reaction and hydrolysis of ATP decreases the affinity of RecA protein to DNA. When ATP is replaced by ATPγS, RecA protein promotes homologous pairing of single-stranded and double-stranded DNA molecules equally well (34), and presynaptic filaments formed in the presence of ATPγS under optimum conditions for homologous pairing resemble those formed in the presence of ATP (35). We found that RecA protein induced TRNOEs of the above oligodeoxyribonucleotides in the presence of ATPγS (Fig. 1*a*), and that the crosspeaks of the transferred NOESY spectra were intense and well resolved (see Fig. 2). Intermolecular crosspeaks between RecA protein and oligodeoxyribonucleotides were not observed, probably because of severely broad signals of RecA protein. On the other hand, in the absence of RecA protein, little NOEs of oligodeoxyribonucleotides were detected (Fig. 1*c*), indicating that the TRNOEs depend on interactions of the oligodeoxyribonucleotides with RecA protein.

The NOEs Are Caused by Specific Binding of DNA to Activated RecA Protein. First, we examined whether the observed interactions between DNA and RecA protein had essential characteristics in common with homologous pairing, specifically a requirement for ATPγS and a preference for DNA over RNA.

Consistent with both the requirement of ATP (or ATPγS) for the formation of active presynaptic filaments and the reduction of affinity for DNA upon hydrolysis of ATP to ADP, TRNOEs of the oligodeoxyribonucleotides induced by the addition of RecA protein were significantly reduced when ATPγS was replaced by ADP (Fig. 1 *b* vs. *a*).

RecA protein binds to RNA with much less affinity than to DNA (36, 37). We have observed that the intensity of TRNOE signals was significantly decreased when DNA was replaced by RNA with the same sequence except for the replacement of U for T (Fig. 1*d*).

These observations indicate that the NOEs observed here are caused by specific interactions of RecA protein with DNA that are the same as those responsible for homologous pairing.

Transferred NOE Analysis of DNA Bound to RecA Protein. The patterns of NOE crosspeaks exhibited by oligodeoxyribonucleotides tested in this study have common features. Unusually intense interresidue crosspeaks between H3′ and H8/H6 were observed in transferred NOESY spectra, whereas by contrast, few, if any, interresidue crosspeaks between H1′ and base protons were detected (Fig. 2*B*). We also observed relatively weak sequential H2′–H8/H6 and H2″–H8/H6 NOEs of comparable intensity (Fig. 2*A*). These are a remarkable contrast to those expected for B-form or A-form DNA (28).

Based on these NOE data, we did structural calculations applying a simulated annealing protocol by use of X-PLOR (26). The final structure for each oligodeoxyribonucleotide was well defined as shown in Fig. 3, the result for d(TACG) as an

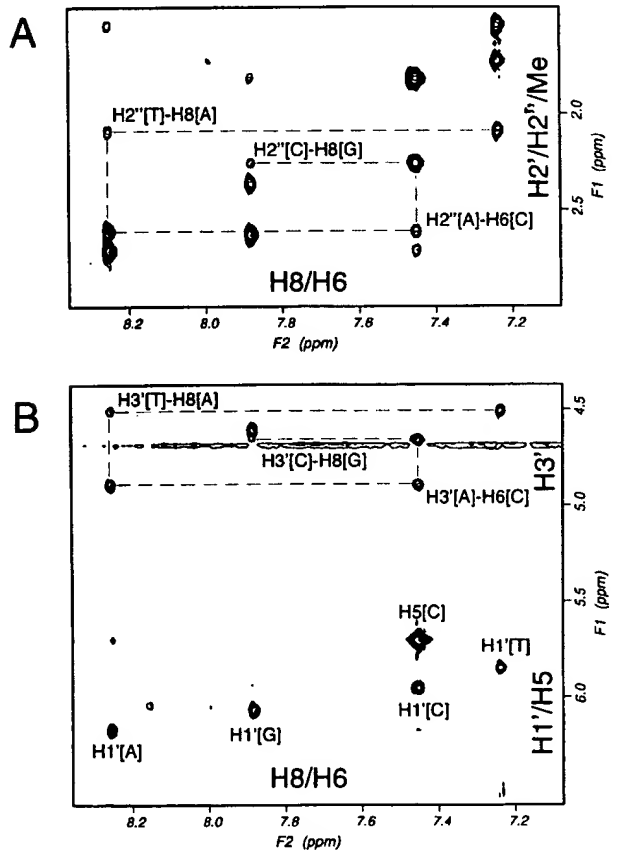


FIG. 2. Two-dimensional transferred NOESY spectra of an oligodeoxyribonucleotide. Two-dimensional transferred NOESY spectra of 0.80 mM d(TACG), 97 μM RecA protein, and 0.80 mM ATPγS at 180 msec mixing time at 25°C. The regions of H8/H6 and H2′/H2″ are shown in *A*, and those of H8/H6 and H3′/H1′ in *B*.

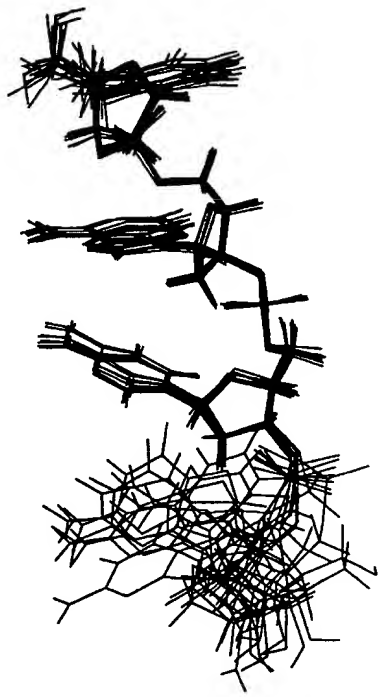


FIG. 3. Superposition of calculated structures of an oligodeoxyribonucleotide, d(TACG). One hundred structures were calculated independently by the use of simulated annealing protocol (X-PLOR; ref. 26). The 10 lowest energy structures are best fitted at the T-A-C region. Total number of NOE constraints is 59; 39 for intraresidue NOEs and 20 for interresidue NOEs. The root-mean-square deviation of the T-A-C region is 0.30 Å. All residues shows similar deoxyribose-base stacking, whereas the fourth residue (G) is disordered because of few NOE constraints due to signal overlapping. There is no violation to the final structure.

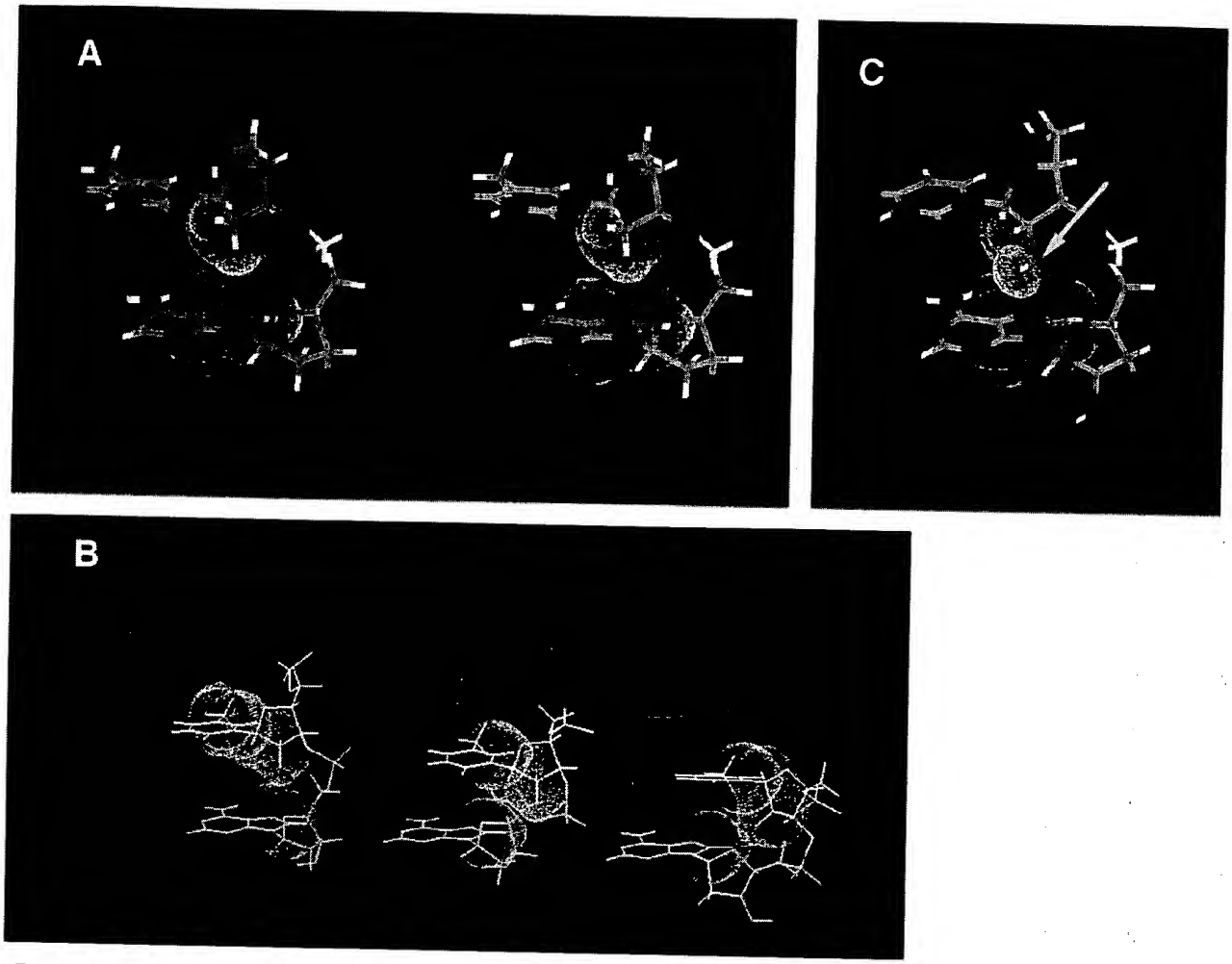


FIG. 4. An extended single-stranded DNA structure induced by RecA protein. (*A*) Stereoview of the representative structure of RecA protein-bound DNA [sequence: d(TA)]. Van der Waals contact surface of a 2′-methylene moiety of a 5′ side residue (T) and those of O4′, C4, C5, N7, C8, H8, and N9 of the 3′ side residue (A) are shown. A structure deduced from the calculation by use of a simulated annealing method was refined using revised parameters as published (27). (*B*) Side views of the structure of DNA in RecA protein-bound form (*Left*), B-form (*Middle*), and A-form (*Right*). Van der Waals contact surfaces between adjacent residues are shown. In B- or A-form DNA the whole structure must be disrupted on the process of strand exchange because of its close packing between adjacent residues. (*C*) A hypothetical RNA structure in the RecA protein-bound form. H2″ is replaced upon a hydroxyl group (indicated by arrow) in RNA.

example. Thus, we conclude that the obtained structure of each oligodeoxyribonucleotide is the major species that mainly contributes to the TRNOE crosspeaks. The calculated structures for all the tested oligodeoxyribonucleotides with variations in sequence and length have a common substructural feature, suggesting that the DNA structure defined in this study is not specific to a sequence or to the size of oligodeoxyribonucleotides.

Fig. 4*A* illustrates a refined molecular model for the structure of single-stranded DNA bound to RecA protein in the presence of ATPγS. If we assume that a helical axis is perpendicular to the base planes as indicated by linear dichroism (38), the axial rise per base is nearly 5 Å (1.5 times that of B-form DNA).

DISCUSSION

By TRNOE analysis, we have determined a three-dimensional structure of single-stranded DNA that has been extended by binding to RecA protein in the presence of ATPγS. The most prominent feature of the DNA structure is in the manner of base stacking. In the normal forms of DNA, adjacent bases are stacked by a van der Waals contact. In contrast, in the RecA protein-bound form, the 2′-methylene moiety of each deoxyribose is located above the base of the next residue in place of the normal base–base stacking, and the bases of the single-stranded DNA are separated by nearly 5 Å (Fig. 4*A*). This spacing agrees well with the 50% extension of single-stranded DNA in presynaptic filaments observed by electron microscopy (Fig. 4*B*; refs. 14 and 39), and the present observations reveal the structural basis for that extension. Interactions between a methylene moiety and an aromatic ring were observed in various biomacromolecules (see ref. 40 for review). There is, to our knowledge, no prior report of extended DNA structures maintained by deoxyribose-base stacking through a methylene-base interaction. On the other hand, another type of deoxyribose-base interaction is found in Z-form DNA (41): the cytidine O4′ oxygen is situated above the six-membered ring of guanine at d(CpG) steps.

What is the meaning of this characteristic deoxyribose-base stacking in the RecA-induced DNA extension? RecA protein has been proposed to bind primarily to the phosphate backbone of single-stranded DNA (42). This type of intermolecular interaction probably triggers the extension of single-stranded DNA upon polymerization of RecA monomers along the DNA backbone. In addition, we propose that the hydrophobic

deoxyribose-base stacking interaction stabilizes intramolecularly the unique DNA conformation. This mechanism presents a striking contrast to that of the widely found DNA extension upon intermolecular stacking interactions, namely, intercalation of aromatic moieties of a dye or amino acid residue between adjacent bases. In this context, it has been suggested that some intramolecular interaction contributes to stabilization of an extended DNA; protein-free DNA molecules, under stress from an external force, undergo a highly cooperative transition into a stretched structure whose length is 1.7 times that of B-form DNA (43, 44)

Judging from the DNA structure revealed by this study, RNA molecules would not form a stable complex with RecA protein, because the 2′-hydroxyl group of RNA will repel the base and the sugar of the following residue (Fig. 4*C*). This would explain previous and current observations that RNA has much less affinity to RecA protein than DNA (Fig. 1*d*).

DNA has an advantage over RNA as material to hold genetic information. A widely accepted reason has been that H2″ confers chemical stabilization on DNA compared with 2′-OH of RNA. Our study suggests another role of H2″ of DNA as genetic material: deoxyribose-base stacking, including 2′-methylene moieties of DNA is required for the binding to RecA protein and its homologs that are general and pivotal machines for homologous recombination. In addition, the deoxyribose-base stacking could be intrinsically required for a homology search between polynucleotides (see below). These newly suggested roles of 2′-methylene moieties might account for the low efficiency and fidelity of homologous recombination in an RNA virus in contrast to high efficiency and accuracy in homologous recombination in organisms with DNA genomes (45).

What is the advantage of the structure stabilized by the deoxyribose-base stacking through a methylene-base interaction? The processes of homologous recognition and strand exchange require rotation of bases so as to exchange partners in base pairs. As described above, RecA protein appears to bind primarily to the phosphate backbone of single-stranded DNA and leaves the bases free for homologous pairing (42). In DNA stabilized by deoxyribose-base stacking, the rotation of adjacent bases is less hindered sterically than in B-form or A-form (Fig. 4*B*). Such freer rotation of bases may favor both homologous pairing and strand exchange.

Another merit of the extended DNA structure is suggested by theoretical conformation analysis of triplex DNA molecules (R-form DNA), which are supposed to be formed during RecA protein-promoted homologous pairing (46–48). According to a structural prediction by the theoretical calculations, the bases of the third strand in the putative triplex DNA would incline and mispair to adjacent base pairs when DNA molecules are not extended (49).

In the presence of ATP or its unhydrolyzable analog, RecA protein binds to double-stranded DNA as well and forms a helical nucleoprotein filament. Double-stranded DNA in the RecA filament has also been found to be extended by 1.5 times as compared with B-form DNA, and to be unwound to 18.6 bp per turn (16, 17). We suppose that RecA-bound double-stranded DNA would be extended by the deoxyribose-base stacking as in the case of single-stranded DNA. We made a model-building study on double-stranded DNA including the deoxyribose-base stacking, and obtained a structure that fits the parameters of RecA filaments (T.N., unpublished work).

Finally, although the structure of a long stretch of DNA in presynaptic filaments could be different from those of the oligodeoxyribonucleotides, we believe that the three-dimensional structure revealed by this study reflects the structure of single-stranded DNA in presynaptic filaments for the following reasons: (*i*) the signals from which the structure was deduced depend on the presence of both RecA protein and ATPγS, (*ii*) the structure can be adopted by DNA but not by RNA, (*iii*) the pattern of NOE crosspeaks is independent of residues in an oligodeoxyribonucleotide and of the sequence and length of the tested oligomers, and (*iv*) the structure agrees well with the extension of single-stranded DNA in the filaments as observed by electron microscopy.

Thus, the structure of the oligodeoxyribonucleotides determined in this study provides a new model that explains how the extended form of DNA in the RecA nucleoprotein filament is stabilized and that further suggests that the functional significance of this form is to facilitate the rotation of bases.

We would like to thank Dr. Yoji Arata (Water Research Institute, Tsukuba) for his suggestions and comments and Dr. Charles M. Radding (Yale Medical School, New Haven) for his suggestions and reading of this manuscript. This study was partly supported by grants for "Biodesign Research Program" from The Institute of Physical and Chemical Research and from the Ministry of Education, Science and Culture, Japan.

1. Clark, A. J. & Margulies, A. D. (1965) *Proc. Natl. Acad. Sci. USA* **53,** 451–459.
2. Kobayashi, I. & Ikeda, H. (1978) *Mol. Gen. Genet.* **166,** 25–29.
3. McEntee, K., Weinstock, G. M. & Lehman, I. R. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 2615–2619.
4. Shibata, T., DasGupta, C., Cunningham, R. P. & Radding, C. M. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 1638–1642.
5. Radding, C. M. (1991) *J. Biol. Chem.* **266,** 5355–5358.
6. Camerini-Otero, R. D. & Hsieh, P. (1993) *Cell* **73,** 217–223.
7. Kowalczykowski, S. C., Dixon, D. A., Eggleston, A. K., Lauder, S. D. & Rehrauer, W. M. (1994) *Microbiol. Rev.* **58,** 401–465.
8. Stasiak, A. & Egelman, E. H. (1994) *Experientia* **50,** 192–203.
9. Cox, M. M. (1995) *J. Biol. Chem.* **270,** 26021–26024.
10. Rao, B. J., Chiu, S. K., Bazemore, L. R., Reddy, G. & Radding, C. M. (1995) *Trends Biochem. Sci.* **20,** 109–113.
11. Kurumizaka, H. & Shibata, T. (1996) *J. Biochem. (Tokyo)* **119,** 216–223.
12. Eggleston, A. K. & West, S. C. (1996) *Trends Genet.* **12,** 20–26.
13. Shibata, T., Cunningham, R. P., DasGupta, C. & Radding, C. M. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 5100–5104.
14. Flory, J., Tsang, S. S. & Muniyappa, K. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 7026–7030.
15. Howard-Flanders, P., West, S. C. & Stasiak, A. (1984) *Nature (London)* **309,** 215–220.
16. Di Capua, E., Engel, A., Stasiak, A. & Koller, T. (1982) *J. Mol. Biol.* **157,** 87–103.
17. Dunn, K., Chrysogelos, S. & Griffith, J. (1982) *Cell* **28,** 757–765.
18. Ogawa, T., Yu, X., Shinohara, A. & Egelman, E. H. (1993) *Science* **259,** 1896–1899.
19. Benson, F. E., Stasiak, A. & West, S. C. (1994) *EMBO J.* **13,** 5764–5771.
20. Yu, X. & Egelman, E. H. (1993) *J. Mol. Biol.* **232,** 1–4.
21. Shibata, T., Cunningham, R. P. & Radding, C. M. (1981) *J. Biol. Chem.* **256,** 7557–7564.
22. Shibata, T., Osber, L. & Radding, C. M. (1983) *Methods Enzymol.* **100,** 197–209.
23. Lowry, O. H., Rosebrough, N. J., Farr, A. L. & Randall, R. J. (1951) *J. Biol. Chem.* **193,** 265–275.
24. Scherf, T. & Anglister, J. (1993) *Biophys. J.* **64,** 754–761.
25. Rance, M., Bodenhausen, G., Wagner, G., Wüthrich, K. & Ernst, R. R. (1985) *J. Magn. Reson.* **62,** 497–510.
26. Brünger, A. T. (1987) *X-PLOR* Manual (Yale Univ. Press, New Haven, CT), Version 3.1.
27. Parkinson, G., Vojtechovsky, J., Clowney, L., Brünger, A. T. & Berman, H. M. (1996) *Acta Crystallogr. D* **52,** 57–64.
28. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids* (Wiley, New York).
29. Clore, G. M. & Gronenborn, A. M. (1982) *J. Magn. Reson.* **48,** 402–417.
30. Clore, G. M. & Gronenborn, A. M. (1983) *J. Magn. Reson.* **53,** 423–442.
31. Landy, S. B. & Rao, B. D. N. (1989) *J. Magn. Reson.* **81,** 371–377.
32. Campbell, A. P. & Sykes, B. D. (1991) *J. Magn. Reson.* **93,** 77–92.
33. Ni, F. & Scheraga, H. A. (1994) *Acc. Chem. Res.* **27,** 257–264.
34. Cox, M. M. & Lehman, I. R. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 3433–3437.

35. Honigberg, S. M., Gonda, D. K., Flory, J. & Radding, C. M. (1985) *J. Biol. Chem.* **260,** 11845–11851.
36. McEntee, K., Weinstock, G. M. & Lehman, I. R. (1981) *J. Biol. Chem.* **256,** 8835–8844.
37. Kirkpatrick, D. P., Rao, B. J. & Radding, C. M. (1992) *Nucleic Acids Res.* **20,** 4339–4346.
38. Norden, B., Elvingson, C., Kubista, M., Sjoberg, B., Ryberg, H., Ryberg, M., Mortensen, K. & Takahashi, M. (1992) *J. Mol. Biol.* **226,** 1175–1191.
39. Stasiak, A. & Di Capua, E. (1982) *Nature (London)* **299,** 185–186.
40. Nishio, M., Umezawa, Y., Hirota, M. & Takeuchi, Y. (1995) *Tetrahedron* **51,** 8665–8701.
41. Egli, M. & Gessner, R. V. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 180–184.
42. Leahy, M. C. & Radding, C. M. (1986) *J. Biol. Chem.* **261,** 6954–6960.
43. Cluzel, P., Lebrun, A., Heller, C., Lavery, R., Viovy, J. L., Chatenay, D. & Caron, F. (1996) *Science* **271,** 792–794.
44. Smith, S. B., Cui, Y. & Bustamante, C. (1996) *Science* **271,** 795–799.
45. Lai, M. M. C. (1992) *Microbiol. Rev.* **56,** 61–79.
46. Hsieh, P., Camerini-Otero, C. S. & Camerini-Otero, R. D. (1990) *Gene Dev.* **4,** 1951–1963.
47. Umlauf, S. W., Cox, M. M. & Inman, R. B. (1990) *J. Biol. Chem.* **265,** 16898–16912.
48. Rao, B. J., Dutreix, M. & Radding, C. M. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 2984–2988.
49. Zhurkin, V. B., Raghunathan, G., Ulyanov, N. B., Camerini-Otero, R. D. & Jernigan, R. L. (1994) *J. Mol. Biol.* **239,** 181–200.

Article No. jmbi.1999.2904 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (1999) 290, 495–504

JMB

The N-terminal Domain of the Human Rad51 Protein Binds DNA: Structure and a DNA Binding Surface as Revealed by NMR

Hideki Aihara[1,2], Yutaka Ito[1,5], Hitoshi Kurumizaka[3,4], Shigeyuki Yokoyama[2,3,4] and Takehiko Shibata[1,5]*

[1]*Cellular & Molecular Biology Laboratory, The Institute of Physical and Chemical Research (RIKEN), 2-1 Hirosawa, Wako-shi, Saitama, 351-0198, Japan*

[2]*Department of Biophysics and Biochemistry, The Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan*

[3]*Cellular Signaling Laboratory and*

[4]*Genomic Sciences Center The Institute of Physical and Chemical Research (RIKEN) 2-1 Hirosawa, Wako-shi Saitama, 351-0198, Japan*

[5]*CREST, JST (Japan Science and Technology)*

**Corresponding author*

Human Rad51 protein (HsRad51) is a homolog of *Escherichia coli* RecA protein, and functions in DNA repair and recombination. In higher eukaryotes, Rad51 protein is essential for cell viability. The N-terminal region of HsRad51 is highly conserved among eukaryotic Rad51 proteins but is absent from RecA, suggesting a Rad51-specific function for this region. Here, we have determined the structure of the N-terminal part of HsRad51 by NMR spectroscopy. The N-terminal region forms a compact domain consisting of five short helices, which shares structural similarity with a domain of endonuclease III, a DNA repair enzyme of *E. coli*. NMR experiments did not support the involvement of the N-terminal domain in HsRad51-HsBrca2 interaction or the self-association of HsRad51 as proposed by previous studies. However, NMR tiration experiments demonstrated a physical interaction of the domain with DNA, and allowed mapping of the DNA binding surface. Mutation analysis showed that the DNA binding surface is essential for double-stranded and single-stranded DNA binding of HsRad51. Our results suggest the presence of a DNA binding site on the outside surface of the HsRad51 filament and provide a possible explanation for the regulation of DNA binding by phosphorylation within the N-terminal domain.

Keywords: Rad51 protein; NMR spectroscopy; solution structure; DNA binding; genetic recombination

Introduction

The human Rad51 protein is a homolog of *Escherichia coli* RecA protein and *Saccharomyces cerevisiae* Rad51 protein (Shinohara *et al.*, 1993). *S. cerevisiae RAD51* gene, along with other members of the *RAD52* epistasis group of genes including *RAD50, RAD52, RAD54, RAD55* and *RAD57*, functions in DNA double-strand break repair and genetic recombination (Petes *et al.*, 1991; Resnick, 1987; Shinohara *et al.*, 1992). Although the precise cellular role of HsRad51 is not fully understood, it is believed to be involved in DNA repair and recombination (see Baumann & West, 1998; Vispe & Defais, 1997). HsRad51 has similar biochemical properties to RecA and yeast Rad51, i.e. it catalyzes *in vitro* the pairing and exchange of homologous double-stranded DNA and single-stranded DNA (Baumann *et al.*, 1996; Gupta, *et al.*, 1997; Sung, 1994). However, the activity of HsRad51 is significantly lower than that of RecA. This suggests a requirement for additional factors in Rad51 functioning, and several biochemical studies have shown that HsRPA (Baumann *et al.*, 1996; Sung, 1994), HsRad52 (Benson *et al.*, 1998; Shen *et al.*, 1996), and HsRad54 (Golub *et al.*, 1997; petukhova *et al.*, 1998) stimulate HsRad51-mediated reactions through direct interactions with HsRad51. Whereas yeast cells deficient in Rad51 are viable (Shinohara

Abbreviations used: HsRad51, the human Rad51 protein; NOE, nuclear Overhauser enhancement; NOESY, nuclear Overhauser enhancement spectroscopy; HSQC, heteronuclear single quantum correlation; TOCSY, total correlation spectroscopy; HMQC, heteronuclear multiple quantum correlation.

E-mail address of the corresponding author: tshibata@postman.riken.go.jp

0022-2836/99/270495–10 $30.00/0

et al., 1992), transgenic mice lacking Rad51 die in the early stage of the development (Sonoda *et al.*, 1998), and chicken B-cells stop cell-growth when the expression of *Rad51* is depressed (Sonoda *et al.*, 1998). In addition, HsRad51 has been found to interact with several tumor suppressors namely p53 (Buchhop *et al.*, 1997; Sturzbecher *et al.*, 1996), Brca1 (Scully *et al.*, 1997) and Brca2 (Chen *et al.*, 1998; Katagiri *et al.*, 1998; Mizuta *et al.*, 1997; Sharan *et al.*, 1997; Wong *et al.*, 1997). These findings suggest essential roles for Rad51 in cell proliferation and genome maintenance in higher eukaryotes.

Alignment of the amino acid sequences of RecA and HsRad51 shows that the central domain of RecA is homologous to the C-terminal portion (approximately two-thirds from the C terminus) of HsRad51 (Shinohara *et al.*, 1993). HsRad51 has extra sequences on its N terminus side, whereas RecA has an extra C-terminal domain which comprises the DNA binding surface (Figure 1; Aihara *et al.*, 1997; Kurumizaka *et al.*, 1996). The N-terminal region (amino acid residues 1-95) of HsRad51 is well conserved among eukaryotic Rad51 proteins, but is absent from RecA. This suggests an important role for this region in Rad51-specific functions such as interactions with other proteins. Indeed, yeast two-hybrid analyses showed that the N-terminal region of Rad51 mediates both Rad51-Rad52 interaction and the self-association of Rad51 in *S. cerevisiae* (Donovan *et al.*, 1994), and a small region near the N terminus (amino acid residues 1-43) of mouse Rad51 protein (MmRad51) is essential for the interaction with MmBrca2 (Sharan *et al.*, 1997). The importance of the N-terminal region is further supported by the recent finding that c-Abl tyrosine kinase regulates HsRad51 function through the phosphorylation on Tyr54 (Yuan *et al.*, 1998). It may also be possible that the N-terminal region of Rad51 takes the place of the C-terminal domain of RecA, and functions in DNA binding.

While the tertiary structure of HsRad51 has not been clarified, the crystal structure of RecA has been determined (Story *et al.*, 1992). An electron microscopic study demonstrated that HsRad51-DNA filaments resembled those of RecA (Benson *et al.*, 1994; Ogawa *et al.*, 1993). This result, combined with the considerable degree of sequence homology between the C-terminal portion of HsRad51 and the core domain of RecA, suggest that the structures of these two proteins are very similar within the homologous region. However, an amino acid sequence homology search has revealed neither the structure nor the function of the N-terminal region of HsRad51. We anticipated that structural information about the N-terminal region might provide a clue about the function of HsRad51.

Here, we describe the structure determination and functional analysis of the N-terminal region of HsRad51. It was found that the N-terminal region folds into a distinct domain with an all-helical fold, and that the domain carries a DNA binding surface. The results are described in detail below, and the possible roles of the N-terminal domain in the homologous pairing reaction are discussed.

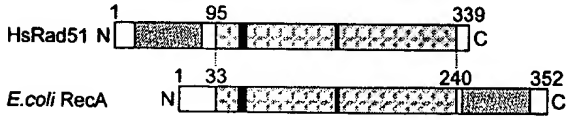


Figure 1. Comparison of HsRad51 and RecA. Amino acid sequences of HsRad51 and *E. coli* RecA are aligned as described (Shinohara *et al.*, 1993). Striped bars indicate the N-terminal domain of HsRad51 identified in this study, and the C-terminal domain of RecA. Shaded bars show conserved regions, with black bars showing the ATP binding consensus sequences.

Results

Structure of the N-terminal region of HsRad51

An N-terminal fragment of HsRad51 containing residues 1 to 114 of the full-length protein, which was found to be highly soluble and monomeric in solution, was used in the NMR study. Residues 1 to 15 and 86 to 114 are disordered as judged by narrower ^{1}H NMR linewidths compared with that of the structured region of the polypeptide, presence of strong HN-H_2O crosspeaks in the ^{15}N-separated NOESY spectrum (Marion *et al.*, 1989), and the absence of long range nuclear Overhauser enhancement (NOE) signals. Thus, the N and C terminal parts were not included in the structure calculation. The solution structure of the segment consisting of residues 16 to 85 was calculated from a total of 1388 NMR-derived restraints (Table 1) using the simulated annealing protocol with the program X-PLOR (Brünger, 1992; Nilges *et al.*, 1988). The backbone (N, C^α, C') superposition for

Table 1. Statistics for the final ensemble of 30 structures

A. *Root mean square deviations from experimental restraints*	
Distance restraints (Å)	
All (1321)	0.029 ± 0.001
Interproton distances	
Intraresidue (285)	0.022 ± 0.003
Sequential (240)	0.022 ± 0.003
Short-range ($2 \leqslant \|i-j\| \leqslant 4$) (183)	0.046 ± 0.003
Long-range ($\|i-j\| \geqslant 5$) (103)	0.057 ± 0.003
Ambiguous (489)	0.012 ± 0.003
Hydrogen bonds (21)	0.045 ± 0.006
Dihedral angle restraints (°) (67)	0.46 ± 0.13
B. *Root mean square deviations from idealized geometry*	
Bonds (Å)	0.00313 ± 0.00013
Angle (°)	0.64 ± 0.01
Improper (°)	0.53 ± 0.02
C. *Coordinate precision for residues 24-79 (Å)*	
Backbone	0.48 ± 0.08
Heavy-atoms	1.06 ± 0.01
D. *PROCHECK[a] Ramachandran map analysis (all structures) (%)*	
Most favoured regions	57.1
Additional allowed regions	33.7
Generously allowed regions	6.7
Disallowed regions	2.5

[a] Laskowski *et al.* (1996).

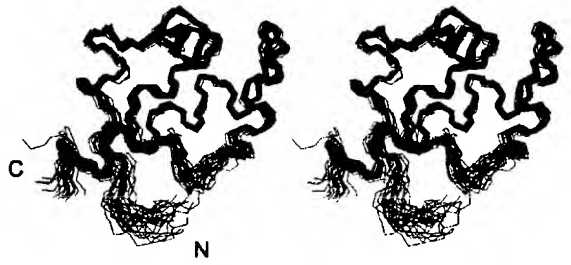


Figure 2. Structure of the N-terminal domain of HsRad51. Stereoview showing the backbone (N, C^{α}, C') atoms of 30 superimposed NMR-derived structures for residues 19-83. This Figure was generated using the program MIDASPlus (Ferrin *et al.*, 1988).

an ensemble of the final 30 structures is presented in Figure 2. An alternate minor backbone conformation was indicated for the region including residues Gly21-Pro22-Gln23 and Val49-Glu50-Ala51, where two sets of signals were observed in the ^{1}H-^{15}N heteronuclear single quantum correlation (HSQC) spectrum (Bodenhausen & Ruben, 1980; Grzesiek & Bax, 1993). Higher mobility may also be present within residues Ala53-Tyr54, whose crosspeaks were missing in the ^{1}H-^{15}N HSQC spectrum. We noted that despite the fact that the domain carries an overall negative charge, a cluster of Lys residues makes a positively charged patch on the protein surface (Figure 3).

Protein interactions

We first investigated whether the N-terminal domain is involved in the protein-protein interactions as suggested by previous studies including those on RecA (Donovan *et al.*, 1994; Sharan *et al.*, 1997; Story *et al.*, 1992). Two polypeptides were tested for their capacity to interact with HsRad51(1-114): a fragment of HsBrca2 that contains residues 3273 to 3309 and the full-length HsRad51 itself. HsBrca2(3273-3309) is 95% identical with MmBrca2(3196-3232), which was identified as the minimal region of MmBrca2 needed for the interaction with the N-terminal region of MmRad51 by the yeast two-hybrid analysis (Sharan *et al.*, 1997). ^{15}N-labeled HsRad51(1-114) was titrated with each of the polypeptides, and the interactions were monitored by measuring a series of ^{1}H-^{15}N HSQC spectra. However, irrespective of the protein used, the spectra did not change throughout the titration. Therefore the NMR experiments do not support the involvement of the N-terminal domain in the interaction with HsBrca2 (3273-3309), or in the self-association of HsRad51. We also found by NMR experiment and GST-pull-down analysis that the N-terminal domain of HsRad51 does not bind HsRad52 neither (H.K., unpublished results).

Interaction with DNA

To test another possible function, we next examined if the N-terminal domain of HsRad51 interacts with DNA using chemical shift perturbation experiments. The titration of a 12 bp double-stranded DNA into the NMR sample of ^{15}N-labeled HsRad51(1-114) caused shifting and broadening of selected crosspeaks in the ^{1}H-^{15}N HSQC spectrum (Figure 4(a)), while the chemical shifts of the remaining residues (including those in the unstructured regions) were only slightly affected or not affected at all. This result indicates a direct interaction between HsRad51(1-114) and DNA. Shifting of the crosspeaks was also observed in a similar titration using 12mer single-stranded DNA, though the chemical shift changes were small. The mode of the chemical shift change indicates that the binding behavior is fast exchange on the NMR timescale, which did not allow us the structure determination of the protein-DNA complex. The dissociation constant (K_d) values were estimated to be 0.31 mM and 0.89 mM in the double-stranded and single-stranded DNA binding, respectively (Figure 5). The resonances affected were those of the backbone amides of Ile61, Lys64, Gly65, Ile66, Ala69, and the neighboring residues on the protein surface (Figure 4(b) and (c)). This indicates that the surface encompassed by these residues furnishes the binding site for DNA. The region overlaps with

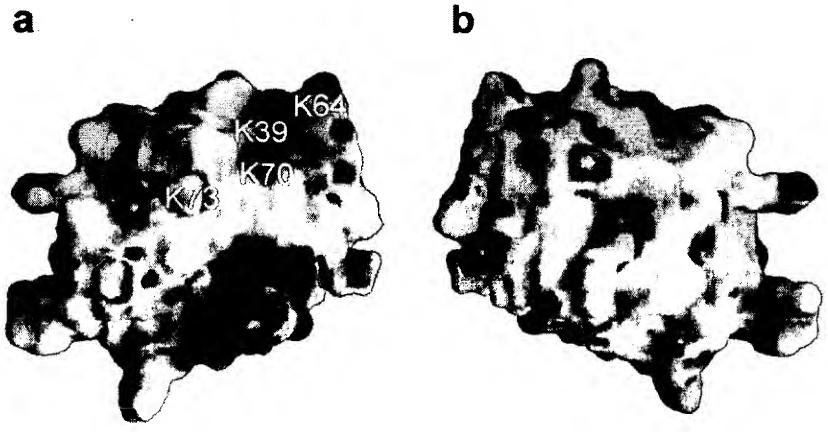
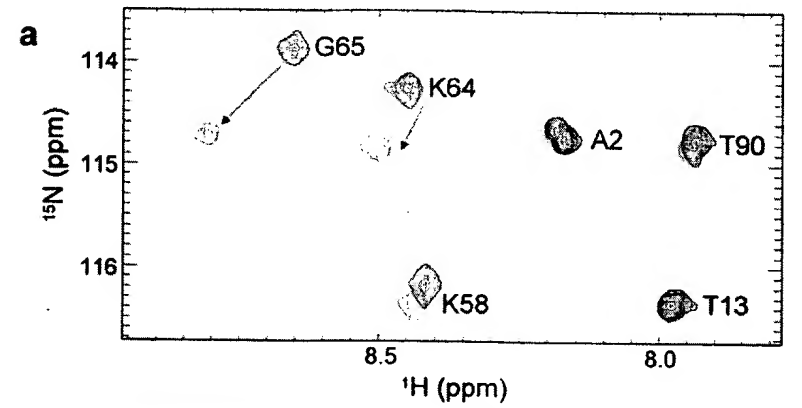


Figure 3. Electrostatic surface potential calculated using the program GRASP (Nicholls & Honig, 1992). Positive potential is colored blue and negative potential is colored red. (a) Same orientation as shown in Figure 2. (b) Viewed after 180° rotation around the vertical axis.



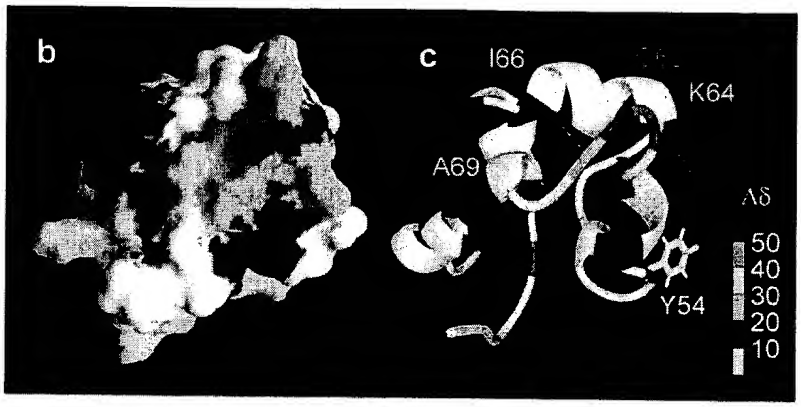


Figure 4. Chemical shift perturbation upon the DNA binding. (a) Expansions of ^{1}H-^{15}N HSQC spectra of ^{15}N-labeled HsRad51(1-114) in the absence (black contours) and presence (red contours) of a three molar equivalent of 12 bp double-stranded DNA. The crosspeaks that shift upon the addition of DNA are indicated. (b), (c) Chemical shift change of backbone ^{1}H and ^{15}N calculated as $[(\Delta\delta^{1}H)^2+ (\Delta\delta^{15}N)^2]^{1/2}$ (Hz) is color-coded and mapped onto the (b) molecular surface or the (c) ribbon drawing of HsRad51(19-83). (b) and (c) Drawn using programs GRASP (Nicholls & Honig, 1992) and MIDASPlus (Ferrin *et al.*, 1988), respectively.

the positively charged patch (Figure 3) mentioned before.

Interestingly, a search of the Brookhaven Protein Data Bank with the program Dali (Holm & Sander, 1993) showed that the structure of the N-terminal domain of HsRad51 is similar to that of the six-helix barrel domain of endonuclease III, a DNA repair enzyme of *E. coli* (Thayer *et al.*, 1995; Figure 6). The six-helix barrel domain of *E. coli* endonuclease III is known to be involved in DNA binding, indicating that the structural similarity between the N-terminal domain of HsRad51 and the six-helix barrel domain is functionally relevant.

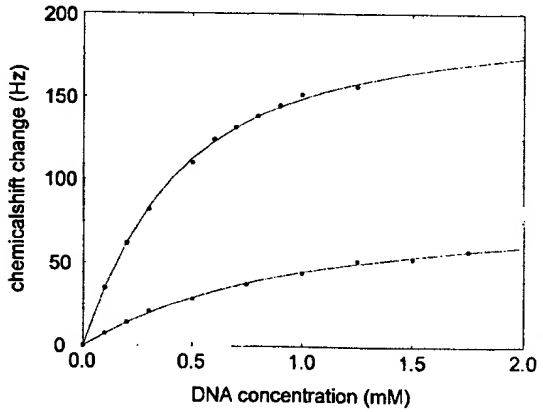


Figure 5. Chemical shift change of backbone ^{1}H and ^{15}N of Gly65 in the titration with 12 bp double-stranded DNA (●) or 12mer single-stranded DNA (■), calculated as $[(\Delta\delta^{1}H)^2+ (\Delta\delta^{15}N)^2]^{1/2}$ (Hz). The continuous line is the best fit of the data to the equation described in Materials and Methods. The concentration of HsRad51(1-114) was 0.2 mM.

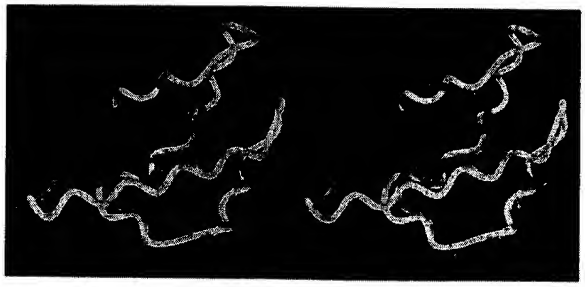


Figure 6. Structural similarity of the N-terminal domain of HsRad51 and the six-helix barrel domain of *E. coli* endonuclease III. Stereodiagram showing the backbone superposition of HsRad51 (red, residues 26-84) and endonuclease III (cyan, residues 31-99) (Thayer *et al.*, 1995). The r.m.s.d. along the C^{α} atoms of residues 26-29, 32-35, 41-45, 46-49, 50-53, 55-65 and 67-84 of HsRad51 with the corresponding part of endonuclease III is 2.86 Å. The Figure was generated using the program MIDASPlus (Ferrin *et al.*, 1988).

However, there may be no evolutionary relationship between these two proteins, since the N-terminal domain of HsRad51 lacks a counterpart for the helix-hairpin-helix DNA binding motif present in the six-helix barrel domain of the endonuclease III.

Mutation analyses

To confirm the functional significance of the DNA binding surface identified by the NMR experiment, we made mutants of HsRad51(1-114) and full-length HsRad51. In order to diminish the positive charge and perturb the local conformation around the DNA binding surface, Lys64 was replaced by a Gly residue (K64G) in both proteins. We first compared the ^{1}H-^{15}N HSQC spectrum of HsRad51(1-114)·K64G with that of the wild-type fragment. A significant difference in the backbone ^{1}H and ^{15}N chemical shift was found for residues within or in the neighborhood of the loop region between Asn62 and Ser67. This was in contrast to the chemical shift of the other residues which did not change or changed only very little (Figure 7(a)). This indicates that the mutation perturbed the local conformation around the DNA binding surface. Chemical shift perturbation upon the addition of DNA was little for HsRad51(1-114)·K64G (Figure 7(b)), suggesting that the DNA binding was diminished by the mutation.

We then examined the DNA binding property of the full-length mutant HsRad51 (K64G) by the gel mobility shift assay. Wild-type HsRad51 makes complexes with single-stranded DNA and double-

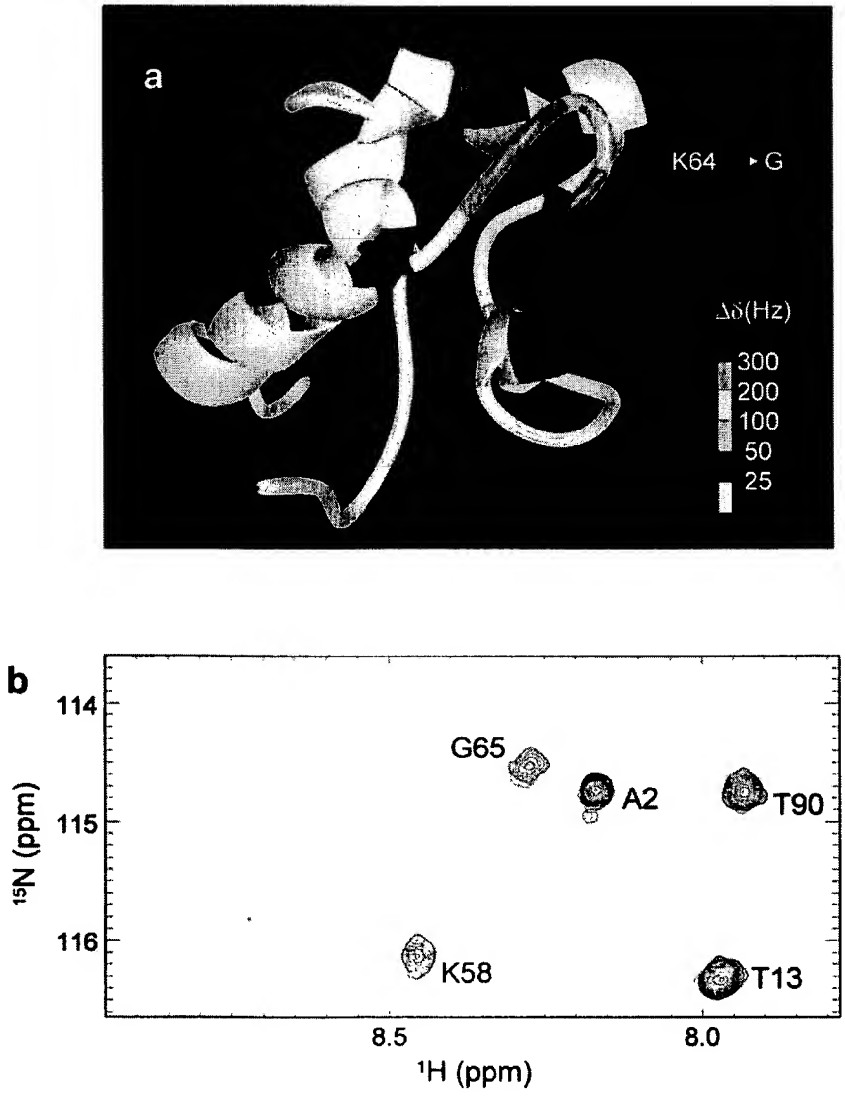


Figure 7. (a) Chemical shift difference between the wild-type and K64G. Differences in the backbone chemical shift calculated as $[(\Delta\delta^{1}H)^2+ (\Delta\delta^{15}N)^2]^{1/2}$ are color-coded. The Figure was drawn using the program MIDASPlus (Ferrin *et al.*, 1988). (b) Expansions of ^{1}H-^{15}N HSQC spectra of ^{15}N-labeled HsRad51(1-114) ·K64G in the absence (black contours) and presence (red contours) of a three molar equivalent of 12 bp double-stranded DNA. Same region as that shown in Figure 4(a).

stranded DNA as shown by the reduced mobility of these DNAs through polyacrylamide or agarose gels (Figure 8). HsRad51·K64G showed decreased single-stranded DNA and double-stranded DNA binding activity compared to the wild-type protein. These results indicate that the DNA binding surface within the N-terminal domain plays an important role in the DNA binding of HsRad51.

Discussion

In the absence of significant amino acid sequence homology to other proteins, we determined the three-dimensional structure of the N-terminal domain of HsRad51 in order to elucidate its function. NMR experiments and a mutation analysis revealed that the N-terminal domain is involved in DNA binding.

Like its bacterial homolog RecA, HsRad51 catalyzes the pairing of single-stranded DNA and double-stranded DNA sharing homologous sequences. It also catalyzes the following strand exchange (Baumann *et al.*, 1996; Baumann & West, 1997). Both RecA and Rad51 bind to single-stranded DNA to form well-conserved right-handed helical filaments, containing single-stranded DNA along their axes (Benson *et al.*, 1994; Ogawa *et al.*, 1993). In the case of RecA, the pairing reaction starts with the formation of a RecA-single-stranded DNA filament, which then incorporates double-stranded DNA (Kahn & Radding, 1984; Shibata *et al.*, 1979; West *et al.*, 1980). In the present study we identified a DNA binding surface within the N-terminal domain of HsRad51. This finding has implications for the function of the N-terminal domain in the homologous pairing reaction. As shown in Figure 1, HsRad51 has an extra segment on its N terminus side situated outside of the homologous region, while RecA has an extra C-terminal domain (Shinohara *et al.*, 1993). In the case of RecA, the C-terminal domain is located on the outside surface of the RecA filament and projects into the helical groove in a pendulous manner (Story *et al.*, 1992; Yu *et al.*, 1998), while the N-terminal part is located on the other side of the groove (Story *et al.*, 1992; Figure 9). In the three-dimensional reconstruction of the electron micrographs of Rad51-DNA filament, the N-terminal domain cannot be visualized due to its flexibility (Ogawa *et al.*, 1993). Assuming similarity between the Rad51 structure and the RecA crystal structure in the homologous region, the N-terminal domain of Rad51 may also be on the outside surface of the Rad51-single-stranded DNA filament and protruding into the helical groove, but from the opposite pole of the protein structure. In our previous bio-

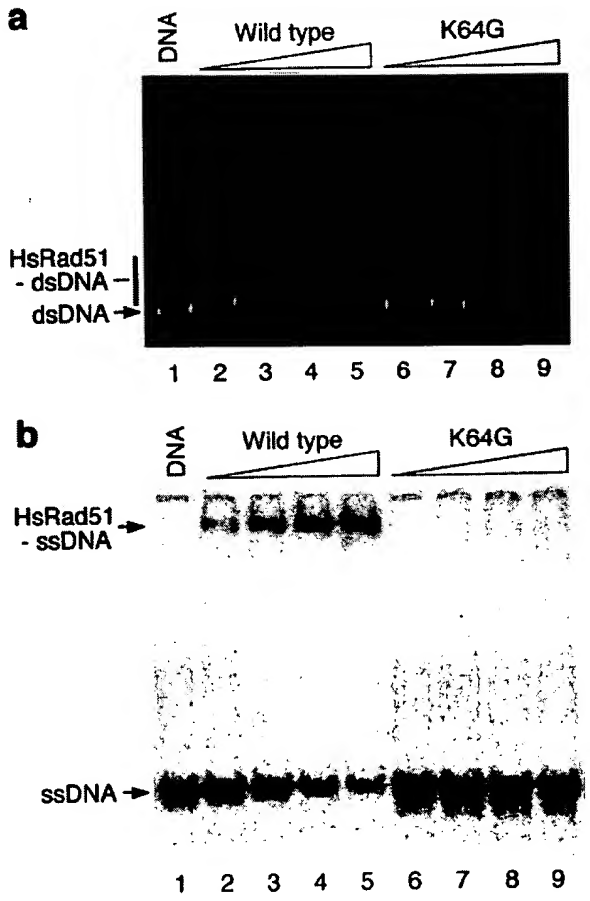


Figure 8. Binding of wild-type and the mutant (K64G) of full-length HsRad51 to double- and single-stranded DNA. (a) Gel mobility shift assay showing the binding of wild-type HsRad51 (lanes 2-5) and HsRad51·K64 G (lanes 6-9) to double-stranded DNA. Protein concentrations were 0 (lane 1), 0.6 μM (lanes 2 and 6), 1.2 μM (lanes 3 and 7), 2.4 μM (lanes 4 and 8), and 4.8 μM (lanes 5 and 9). (b) Gel mobility shift assay showing the binding of wild-type HsRad51 (lanes 2-5) and HsRad51·K64G (lanes 6-9) to single-stranded DNA. Protein concentrations were 0 (lane 1), 0.3 μM (lanes 2,6), 0.6 μM (lanes 3,7), 0.9 μM (lanes 4,8), and 1.2 μM (lanes 5,9).

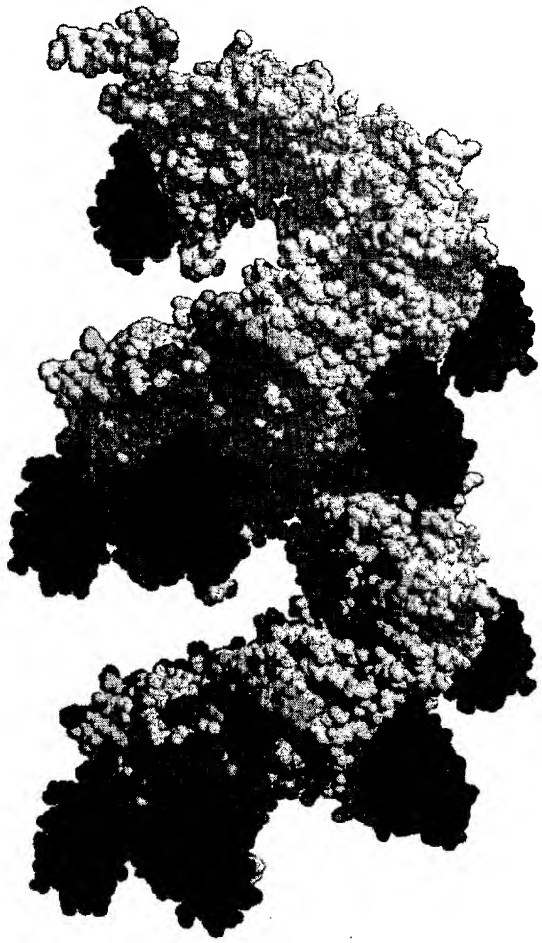


Figure 9. Van der Waals surface of the 12 RecA monomers in the crystal structure (Story *et al.*, 1992). The N-terminal part is shown in cyan and the C-terminal domain in red. The Figure was drawn with MIDAS-Plus (Ferrin *et al.*, 1988).

chemical and NMR analyses we suggested that the DNA binding surface in the C-terminal domain of RecA facilitates the spooling of double-stranded DNA into the helical groove of the RecA-single-stranded DNA filament (Aihara *et al.*, 1997; Kurumizaka *et al.*, 1996). Considering the spatial relationship, the N-terminal domain of HsRad51 could have the same function as the C-terminal domain of RecA, though these two domains share no sequence or structural homology.

From our mutation analysis, it seems likely that the N-terminal DNA binding surface is also important for the binding of HsRad51 to single-stranded DNA. It has been reported that the c-Abl tyrosine kinase inhibits the single-stranded DNA binding of HsRad51 through the phosphorylation on Tyr54 (Yuan *et al.*, 1998). In accordance with this observation, the aromatic side-chain of Tyr54 is not buried inside the molecule (Figure 4(c)). The inhibition of DNA binding by the phosphorylation could be an effect of either electrostatic repulsion or perturbation of the N-terminal domain structure in the DNA binding site identified by this study.

In conclusion, we found that the N-terminal part of the human Rad51 protein constitutes a DNA binding domain. The domain may lie on the outside surface of the Rad51 filament, suggesting the critical role of this domain in the homologous pairing reaction.

Materials and Methods

Sample preparation

HsRad51(1-114) was expressed in *E. coli* strain JM109(DE3) using the pET3a vector (Novagen). $tRNA^{arg3}$ and $tRNA^{arg4}$ were coexpressed to facilitate the translation of the minor codons. Uniformly $^{13}C/^{15}N$ and ^{15}N-labeled proteins were prepared by growing the bacteria on minimal medium containing $^{15}NH_4Cl$ either with or without $^{13}C_6$-glucose. The protein was purified from cell-free extract by successive DEAE-Sepharose Fast Flow (Pharmacia) and Sephadex G50 Superfine (Pharmacia) column chromatography. Mass spectrometry suggested that the N terminus of the protein is acetylated after cleavage of the first methionine residue. Typical NMR samples for structure determination contained 1 mM protein, 20 mM sodium phosphate (pH 6.5), 100 mM NaCl, 2 mM DTT and 0.02% NaN_3 in $^1H_2O/^2H_2O$ (9:1) or ^2H_2O.

Preparation of wild-type and the mutant form of full-length HsRad51 will be described elsewhere (H.K. *et al.*, unpublished results). HsBrca2(3273-3309) was expressed as a glutathione *S*-transferase fusion and purified by affinity chromatography, followed by cleavage with thrombin and cation exchange chromatography.

Structure determination

All NMR spectra were acquired at 30 °C on a Bruker DRX600 or ARX400 spectrometer. The ^1H, ^{13}C and ^{15}N resonances of the backbone were assigned using 3D CBCA(CO)NNH (Grzesiek & Bax, 1992), HNCACB (Wittekine & Mueller, 1993) and 2D ^1H-^{15}N HSQC (Bodenhausen & Ruben, 1980; Grzesiek & Bax, 1993) experiments. The side-chain signals were assigned using 3D HCCH-TOCSY (Bax *et al.*, 1990), H(CCCO)NNH (Grzesiek *et al.*, 1993; Clowes *et al.*, 1993), C(CCO)NNH (Grzesiek *et al.*, 1993; Clowes *et al.*, 1993), ^{15}N-separated TOCSY (Marion *et al.*, 1989), HNHB (Archer *et al.*, 1991), and 2D ^1H-^{13}C HSQC experiments (Bodenhausen & Ruben, 1980). Complete chemical shift assignments have been deposited in the BioMagResBank, with the accession number 4328. Distance restraints were obtained from 3D ^{15}N- or ^{13}C-separated NOESY (Marion *et al.*, 1989) and 2D ^1H-^1H NOESY (Kumar *et al.*, 1980) (recorded in ^2H_2O) spectra with mixing times of 150 ms. NOEs were classified into four distance ranges, 1.8 to 3.2, 1.8 to 3.8, 1.8 to 5.5, and 1.8 to 6.0 Å according to the peak intensities. An additional 0.5 Å was added to the upper limits for distances involving methyl protons. NOE cross-peaks that cannot be assigned unambiguously were included as ambiguous distance restraints (Nilges *et al.*, 1997), restraining the r^{-6} sum of the distances between contributing atoms. The ϕ angle restraints were obtained from the $^3J_{HN,H\alpha}$ coupling constants measured

in HMQC-J (Kay & Bax, 1990) experiment. Hydrogen bond distance restraints and ψ angle restraints were employed for α-helical regions based on the $^3J_{HN,H\alpha}$ coupling constants and chemical shift index (Wishart & Sykes, 1994). Structures were calculated with the random simulated annealing protocol using the program X-PLOR 3.851 (Brünger, 1992; Nilges *et al.*, 1988). The level of ambiguity in the ambiguous distance restraints was reduced during the structure calculation by discarding potential assignments that gave distances >8 Å in the calculated structures with lower target function values. This cut-off was set to 0.5 Å above the upper limit of each distance restraint in the latter stages of the structure calculation. The final structure calculation employed 1300 inter-proton distance restraints, 40 ϕ angle restraints, 21 hydrogen bond distance restraints and 27 ψ angle restraints. Of 100 structures calculated, 70 structures had no restraint violations above 0.5 Å or 5°. The 30 structures with lowest energy were used to represent the solution structure of HsRad51(1-114).

Titration experiments

DNA used in the titration experiment was a 12 bp d(CCGGTGATAGAC)/(GTCTATCACCGG) oligonucleotide, the central ten base-pairs of which was reported to adopt the *B*-type conformation in solution (Baleja *et al.*, 1990). When single-stranded DNA was used, the bottom strand was employed. The DNAs were added to 0.1 or 0.2 mM solution of ^{15}N-labeled HsRad51(1-114), and a series of ^{1}H-^{15}N HSQC spectra were recorded at various single-stranded or double-stranded DNA concentrations ranging from 0.05 mM to 1.75 mM. Dissociation constant (K_d) values were calculated by fitting the experimental data (chemical shift change of Gly65 upon DNA binding) to the equation:

$$K_d = ([P_0] - [PD])([D_0] - [PD])/[PD]$$

$$[PD] = [P_0] \times \Delta_{obs}/\Delta_{max}$$

where $[P_0]$, $[D_0]$, and [PD] are the concentrations of HsRad51(1-114) (total), DNA (total), and the protein-DNA complex, respectively; Δ_{obs} is the difference between the observed chemical shift and the chemical shift of the free state; and Δ_{max} is the chemical shift difference between the free and the bound states. K_d and Δ_{max} were treated as fitting parameters during the curve fitting. In the titration with the full-length HsRad51 and HsBrca2(3273-3309), the unlabeled peptides were added to 0.1 mM solution of ^{15}N- labeled HsRad51(1-114) to give the final concentration of 0.1 mM and 0.2 mM, respectively. All titration experiments were performed at 293 K, and the solution conditions were same as that in the measurements for structure determination.

DNA binding assays

Linearized pGsat4 double-stranded DNA (18 μM, 3216 bp) was incubated with wild-type or the mutant HsRad51 in the 10μl of reaction buffer containing 50 mM Hepes-KOH (pH 7.5), 2 mM ATP, 20 mM creatine phosphate, 1 mM DTT, 100 μg/ml bovine serum albumin, 12 units/ml creatine phosphokinase, 15 mM $MgCl_2$, and 3% (v/v) glycerol. Samples were analyzed by electrophoresis through 0.8% agarose gel in 0.5 × TBE buffer. DNA and DNA-protein complexes were visualized by ethidium bromide staining. In the single-stranded DNA binding assay, ^{32}P-labeled single-stranded oligonucleotide 50 mer (300 nM) was incubated with HsRad51 in the same reaction buffer used in the double-stranded DNA binding assay, except that glycerol content was 6% (v/v). Samples were analyzed by electrophoresis through non-denaturing 12% polyacrylamide gel in TBE buffer. Bands were analyzed by BAS-2500 image analyzer (Fuji).

Protein Data Bank accession numbers

The structure and restraint files have been deposited in the Brookhaven Protein Databank with the accession codes 1b22 and r1b22mr, respectively.

Acknowledgments

We thank Dr N. Dohmae and Dr K. Takio (Division of Biomolecular Characterization, RIKEN) for mass spectrometry, Dr T. Terada (Division of Biomolecular Characterization, RIKEN; Department of Biophysics and Biochemistry, The Graduate School of Science, The University of Tokyo) for helpful discussions. This work was supported by grants from the Biodesign Research Program, and MR Science Program from RIKEN, a grant from the Ministry of Education, Science and Culture, Japan, a research grant from Human Frontier Science Program (RG493/95), and a grant for CREST from JST (Japan Science and Technology) (Y.I. and T.S.).

References

Aihara, H., Ito, Y., Kurumizaka, H., Terada, T., Yokoyama, S. & Shibata, T. (1997). An interaction between a specified surface of the C-terminal domain of RecA protein and double-stranded DNA for homologous pairing. *J. Mol. Biol.* **274**, 213-221.

Archer, S. J., Ikura, M., Torchia, D. A. & Bax, A. (1991). An alternative 3D NMR technique for correlating backbone ^{15}N with side chain Hβ resonances in larger proteins. *J. Magn. Reson.* **95**, 636-641.

Baleja, J. D., Pon, R. T. & Sykes, B. D. (1990). Solution structure of phage λ half-operator DNA by use of NMR, restrained molecular dynamics, and NOE-based refinement. *Biochemistry*, **29**, 4828-4839.

Baumann, P. & West, S. C. (1997). The human Rad51 protein: polarity of strand transfer and stimulation by hRP-A. *EMBO J.* **16**, 5198-5206.

Baumann, P. & West, S. C. (1998). Role of the human RAD51 protein in homologous recombination and double- stranded-break repair. *Trends Biochem. Sci.* **23**, 247-251.

Baumann, P., Benson, F. E. & West, S. C. (1996). Human rad51 protein promotes ATP-dependent homologous pairing and strand transfer reactions *in vitro*. *Cell*, **87**, 757-766.

Bax, A., Clore, G. M. & Gronenborn, A. M. (1990). ^{1}H-^{1}H correlation *via* isotropic mixing of magnetization, a new three-dimensional approach for assigning ^{1}H and ^{13}C spectra of ^{13}C-enriched proteins. *J. Magn. Reson.* **88**, 425-431.

Benson, F. E., Stasiak, A. & West, S. C. (1994). Purification and characterization of the human Rad51 protein, an analogue of *E. coli* RecA. *EMBO J.* **13**, 5764-5771.

Benson, F. E., Baumann, P. & West, S. C. (1998). Synergistic actions of Rad51 and Rad52 in recombination and DNA repair. *Nature*, **391**, 401-404.

Bodenhausen, G. & Ruben, D. J. (1980). Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Letters*, **69**, 185-189.

Brünger, A. T. (1992). *X-PLOR version 3.1. A system for X-ray Crystallography and NMR*, Yale Unversity Press, New Haven and London.

Buchhop, S., Gibson, M. K., Wang, X. W., Wagner, P., Sturzbecher, H. W. & Harris, C. C. (1997). Interaction of p53 with the human Rad51 protein. *Nucl. Acids Res.* **25**, 3868-3874.

Chen, P. L., Chen, C. F., Chen, Y., Xiao, J., Sharp, Z. D. & Lee, W. H. (1998). The BRC repeats in BRCA2 are critical for RAD51 binding and resistance to methyl methanesulfonate treatment. *Proc. Natl Acad. Sci. USA*, **95**, 5287-5292.

Clowes, R. T., Boucher, W., Hardman, C. H., Domaille, P. J. & Laue, E. D. (1993). A 4D HCC(CO)NNH experiment for the correlation of aliphatic side-chain and backbone resonances in $^{13}C/^{15}N$-labelled proteins. *J. Biomol. NMR*, **3**, 349-354.

Donovan, J. W., Milne, G. T. & Weaver, D. T. (1994). Homotypic and heterotypic protein associations control rad51 function in double-strand break repair. *Gene Dev.* **8**, 2552-2562.

Ferrin, T. E., Huang, C. C., Jarvis, E. & Langridge, R. (1988). The MIDAS display system. *J. Mol. Graph.* **6**, 13-27.

Golub, E. I., Kovalenko, O. V., Gupta, R. C., Ward, D. C. & Radding, C. M. (1997). Interaction of human recombination proteins Rad51 and Rad54. *Nucl. Acids Res.* **25**, 4106-4110.

Grzesiek, S. & Bax, A. (1992). Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc.* **114**, 6291-6293.

Grzesiek, S. & Bax, A. (1993). The importance of not saturating H_2O in protein NMR. Application to sensitivity enhancement and NOE measurements. *J. Am. Chem. Soc.* **115**, 12593-12594.

Grzesiek, S., Anglister, J. & Bax, A. (1993). Correlation of backbone amide and aliphatic side-chain resonances in $^{13}C/^{15}N$-enriched proteins by isotropic mixing of ^{13}C magnetization. *J. Magn. Reson.* **101**, 114-119.

Gupta, R. C., Bazemore, L. R., Golub, E. I. & Radding, C. M. (1997). Activities of human recombination protein Rad51. *Proc. Natl Acad. Sci. USA* **94**, 463-468.

Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123-138.

Kahn, R. & Radding, C. M. (1984). Separation of the presynaptic and synaptic phases of homologous pairing promoted by recA protein. *J. Biol. Chem.* **259**, 7495-7503.

Katagiri, T., Saito, H., Shinohara, A., Ogawa, H., Kamada, N., Nakamura, Y. & Miki, Y. (1998). Multiple possible sites of BRCA2 interacting with DNA repair protein RAD51. *Genes Chrom. Cancer*, **21**, 217-222.

Kay, L. E. & Bax, A. (1990). New methods for the measurement of NH-CαH coupling constants in ^{15}N-labeled proteins. *J. Magn. Reson.* **86**, 110-126.

Kumar, A., Ernst, R. R. & Wuthrich, K. (1980). A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochem. Biophys. Res. Commun.* **95**, 1-6.

Kurumizaka, H., Aihara, H., Ikawa, S., Kashima, T., Bazemore, L. R., Kawasaki, K., Sarai, A., Radding, C. M. & Shibata, T. (1996). A possible role of the C-terminal domain of the RecA protein: a gateway model for double-stranded DNA-binding. *J. Biol. Chem.* **271**, 33515-33524.

Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R. & Thornton, J. M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR*, **8**, 477-486.

Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A. & Bax, A. (1989). Three-dimensional heteronuclear NMR of $_{15}N$-labeled proteins. *J. Am. Chem. Soc.* **111**, 1515-1517.

Mizuta, R., LaSalle, J. M., Cheng, H. L., Shinohara, A., Ogawa, H., Copeland, N., Jenkins, N. A., Lalande, M. & Alt, F. W. (1997). RAB22 and RAB163/mouse BRCA2: proteins that specifically interact with the RAD51 protein. *Proc. Natl Acad. Sci. USA*, **94**, 6927-6932.

Nicholls, A. & Honig, B. (1992). *GRASP: Graphical Representation and Analysis of Surface Properties*, Columbia University, New York.

Nilges, M., Clore, G. M. & Gronenborn, A. M. (1988). Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. *FEBS Letters*, **229**, 317-324.

Nilges, M., Macias, M. J., O'Donoghue, S. I. & Oschkinat, H. (1997). Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from beta-spectrin. *J. Mol. Biol.* **269**, 408-422.

Ogawa, T., Yu, X., Shinohara, A. & Egelman, E. H. (1993). Similarity of the yeast RAD51 filament to the bacterial RecA filament. *Science*, **259**, 1896-1899.

Petes, T. D., Malone, R. E. & Symington, L. S. (1991). Recombination in yeast. In *Genome Dynamics, Protein Synthesis, and Energetics. The Molecular and Cellular Biology of the Yeast Saccharomyces* (Broach, J. R., Pringle, J. R. & Jones, E. W., eds), vol. 1, pp. 407-521, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Petukhova, G., Stratton, S. & Sung, P. (1998). Catalysis of homologous DNA pairing by yeast Rad51 and Rad54 proteins. *Nature*, **393**, 91-94.

Resnick, M. A. (1987). Investigating the genetic control of biochemical events in meiotic recombination. In *Meiosis*, Academic Press, London.

Scully, R., Chen, J. J., Plug, A., Xiao, Y. H., Weaver, D., Feunteun, J., Ashley, T. & Livingston, D. M. (1997). Association of BRCA1 with Rad51 in mitotic and meiotic cells. *Cell*, **88**, 265-275.

Sharan, S. K., Morimatsu, M., Albrecht, U., Lim, D. S., Regel, E., Dinh, C., Sands, A., Eichele, G., Hasty, P. & Bradley, A. (1997). Embryonic lethality and radiation hypersensitivity mediated by Rad51 in mice lacking Brca2. *Nature*, **386**, 804-810.

Shen, Z., Cloud, K. G., Chen, D. J. & Park, M. S. (1996). Specific interactions between the human RAD51 and RAD52 proteins. *J. Biol. Chem.* **271**, 148-152.

Shibata, T., Cunningham, R. P., DasGupta, C. & Radding, C. M. (1979). Homologous pairing in gen-

etic recombination: complexes of recA protein and DNA. *Proc. Natl Acad. Sci. USA*, **76**, 5100-5104.

Shinohara, A., Ogawa, H. & Ogawa, T. (1992). RAD51 protein involved in repair and recombination in *S. cerevisiae* is a recA-like protein. *Cell*, **69**, 457-470.

Shinohara, A., Ogawa, H., Matsuda, Y., Ushio, N., Ikeo, K. & Ogawa, T. (1993). Cloning of human, mouse and fission yeast recombination genes homologous to RAD51 and recA. *Nature Genet.* **4**, 239-243.

Sonoda, E., Sasaki, M. S., Buerstedde, J.-M., Bezzubova, O., Shinohara, A., Ogawa, H., Takata, M., Yamaguchi-Iwai, Y. & Takeda, S. (1998). Rad51 deficient vertebrate cells accumulate chromosomal breaks prior to cell death. *EMBO J.* **17**, 598-608.

Story, R. M., Weber, I. & Steitz, T. A. (1992). The structure of the *E. coli recA* protein monomer and polymer. *Nature*, **355**, 318-325.

Sturzbecher, H. W., Donzelmann, B., Henning, W., Knippschild, U. & Buchhop, S. (1996). p53 is linked directly to homologous recombination processes *via* RAD51/RecA protein interaction. *EMBO J.* **15**, 1992-2002.

Sung, P. (1994). Catalysis of ATP-dependent homologous DNA pairing and strand exchange by yeast RAD51 protein. *Science*, **265**, 1241-1243.

Thayer, M. M., Ahern, H., Xing, D. X., Cunningham, R. P. & Tainer, J. A. (1995). Novel DNA binding motifs in the DNA repair enzyme endonuclease III crystal structure. *EMBO J.* **14**, 4108-4120.

Tsuzuki, T., Fujii, Y., Sakumi, K., Tominaga, Y., Nakao, K., Sekiguchi, M., Matsushiro, A., Yoshimura, Y. & Morita, T. (1996). Targeted disruption of the rad51 gene leads to lethality in embryonic mice. *Proc. Natl Acad. Sci. USA*, **93**, 6236-6240.

Vispe, S. & Defais, M. (1997). Mammalian Rad51 protein: a RecA homologue with pleiotropic functions. *Biochimie*, **79**, 587-592.

West, S. C., Cassuto, E., Mursalim, J. & Howard-Flanders, P. (1980). Recognition of duplex DNA containing single-stranded regions by recA protein. *Proc. Natl Acad. Sci. USA*, **77**, 2569-2573.

Wishart, D. S. & Sykes, B. D. (1994). The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *J. Biomol. NMR*, **4**, 171-180.

Wittekind, M. & Mueller, L. (1993). HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-protein and nitrogen resonances with the alpha-and beta-carbon resonances in proteins. *J. Magn. Reson. ser. B*, **101**, 201-205.

Wong, A. K. C., Pero, R., Ormonde, P. A., Tavtigian, S. V. & Bartel, P. L. (1997). RAD51 interacts with the evolutionarily conserved BRC motifs in the human breast cancer susceptibility gene *brca2*. *J. Biol. Chem.* **272**, 31941-31944.

Yu, X., Shibata, T. & Egelman, E. H. (1998). Identification of a defined epitope on the surface of the active RecA-DNA filament using a monoclonal antibody and three-dimensional reconstruction. *J. Mol. Biol.* **283**, 985-992.

Yuan, Z. M., Huang, Y., Ishiko, T., Nakada, S., Utsugisawa, T., Kharbanda, S., Wang, R., Sung, P., Shinohara, A., Weichselbaum, R. & Kufe, D. (1998). Regulation of Rad51 function by c-Abl in response to DNA damage. *J. Biol. Chem.* **273**, 3799-3802.

Edited by P. E. Wright

(Received 12 February 1999; received in revised form 18 May 1999; accepted 19 May 1999)

http://www.academicpress.com/jmb

Supplementary material comprising one Figure and one Table is available from JMB Online

APPENDIX D

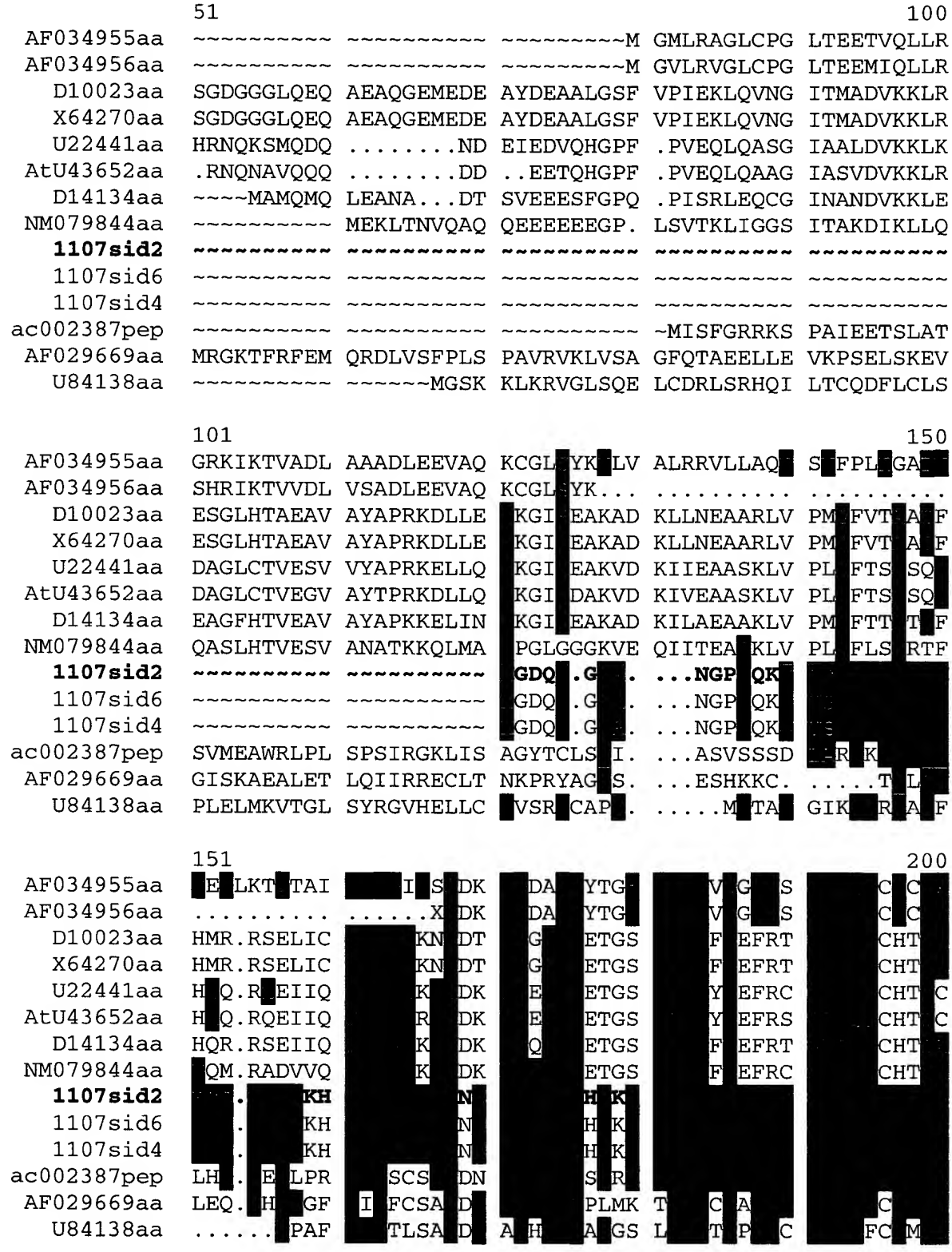
!!AA_MULTIPLE_ALIGNMENT 1.0 PileUp GCG
Attorney Docket: 1107
Serial No.: 09/537,654

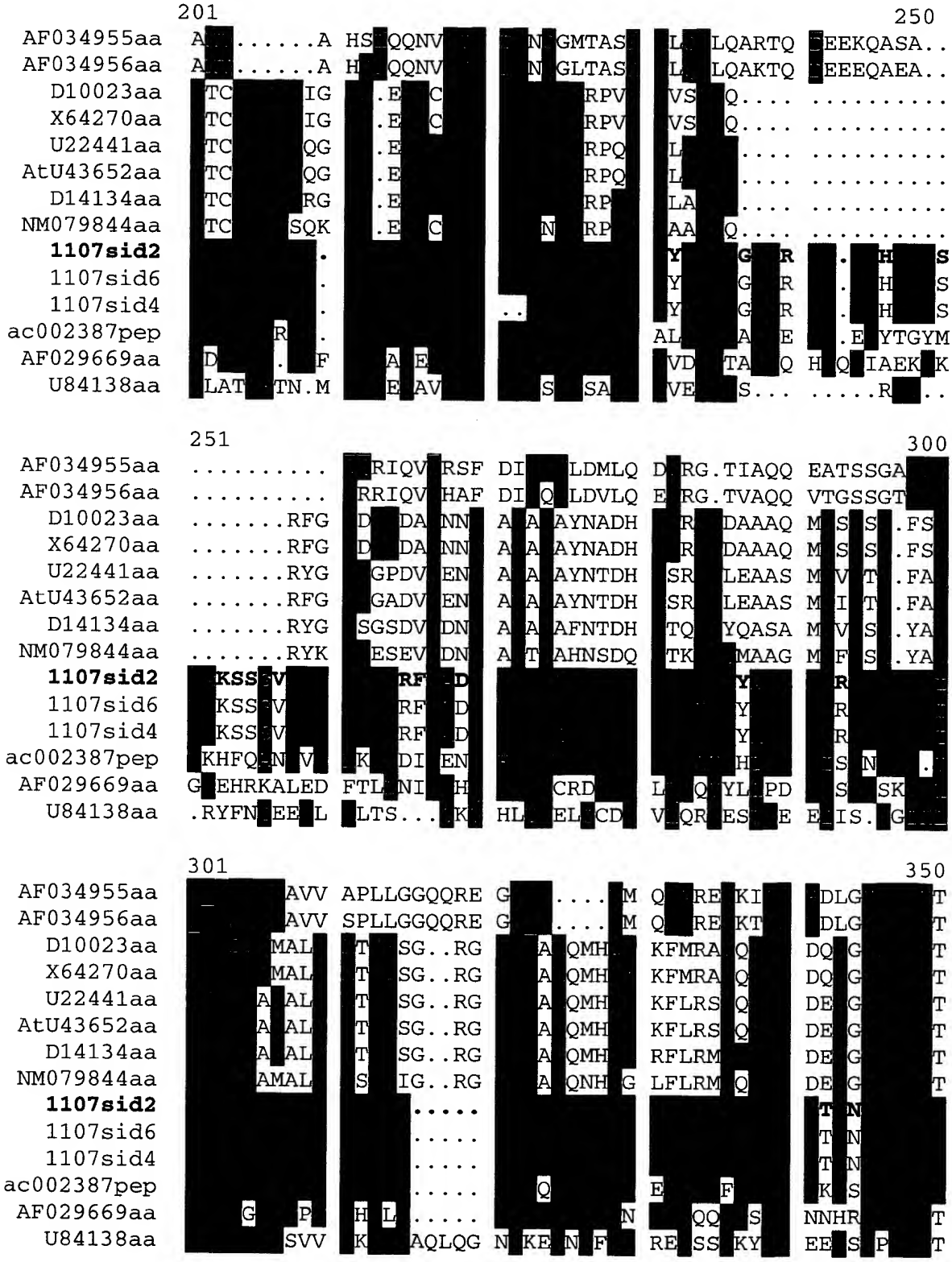
Symbol comparison table: genrundata:blosum62.cmp CompCheck: 1102
GapWeight: 8 GapLengthWeight: 2

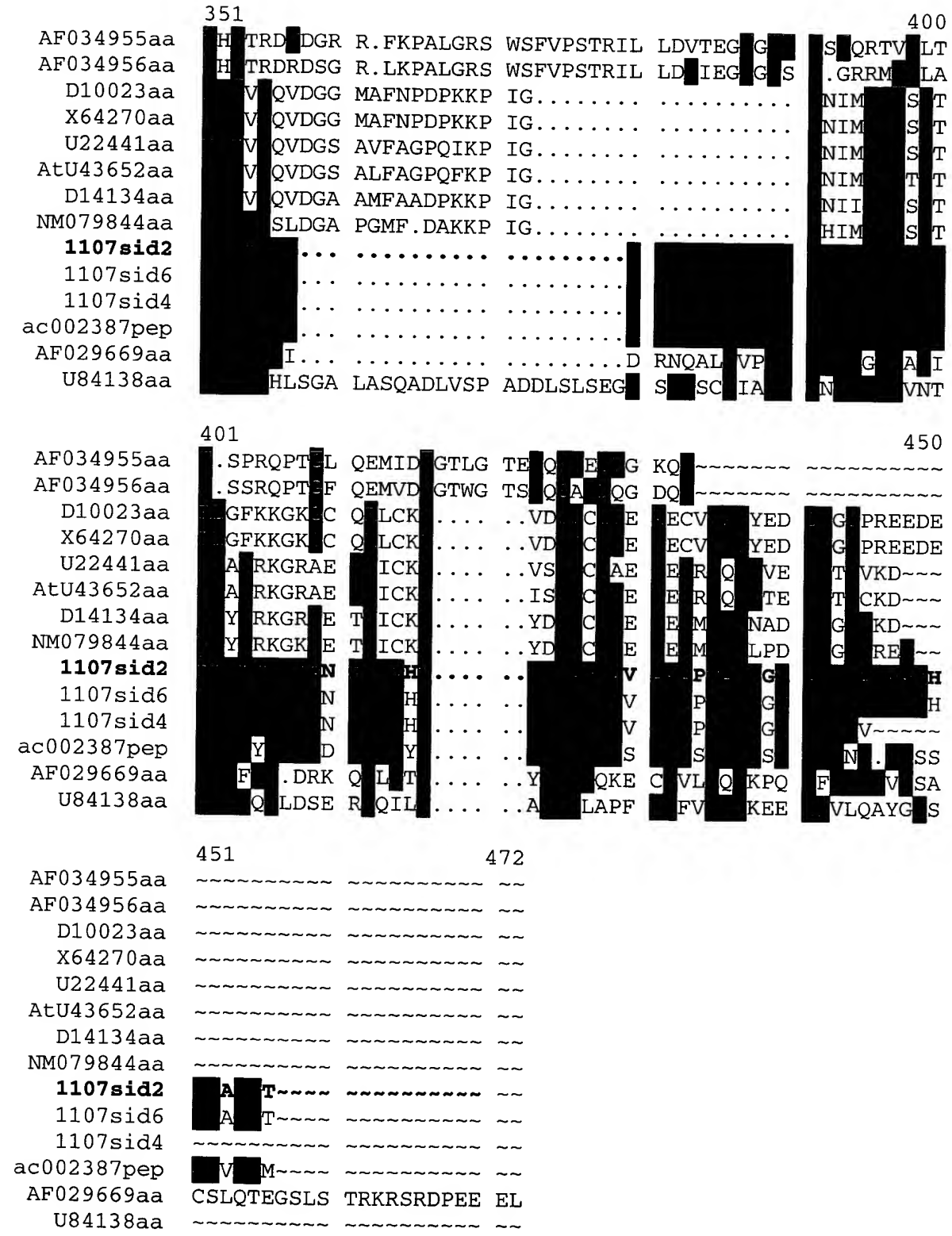
AF034955	Mouse Rad51D protein encoded by AF034955
AF034956	Human Rad51D protein encoded by AF034956
D10023	S. cerevisiae Rad51 protein encoded by D10023
X64270	S. cerevisiae Rad51 protein encoded by X64270
U22441	Tomato Rad51 protein encoded by U22441
U43652	A. thaliani Rad51 protein encoded by U43652
D14134	Human Rad51 protein encoded by D14134
NM079844	D. melanogaster Rad51 protein encoded by NM_079844
1107sid2	**SEQ ID NO: 2 encoded by SEQ ID NO: 1**
1107sid6	SEQ ID NO: 6
1107sid4	SEQ ID NO: 4
ac002387	A. thaliani Rad51C protein encoded by AC002387 At2g45280
AF029669	Human Rad51C protein encoded by AF029669
U84138	Human Rad51B protein encoded by U84138

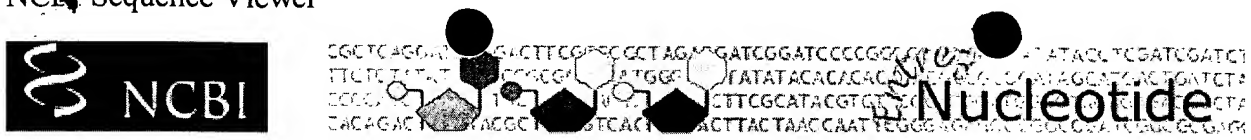
FORMATTING: Amino acid residue [redacted], relative to SEQ ID NO: 2
Amino acid residue [redacted] (conserved substitution)
SEQ ID NO: 2 protein encoded by elected sequence

```
                1                                                   50
AF034955aa      ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
AF034956aa      ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
  D10023aa      MSQVQEQHIS ESQLQYGNGS LMSTVPADLS QSVVDGNGNG SSEDIEATNG
  X64270aa      MSQVQEQHIS ESQLQYGNGS LMSTVPADLS QSVVDGNGNG SSEDIEATNG
   U22441aa     ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~MEQQ
 AtU43652aa     ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~MTTMEQR
  D14134aa      ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
NM079844aa      ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
  1107sid2      ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
  1107sid6      ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
  1107sid4      ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
ac002387pep     ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
 AF029669aa     ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
  U84138aa      ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
```


51
100
AF034955aa ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~M GMLRAGLCPG LTEETVQLLR
AF034956aa ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~M GVLRVGLCPG LTEEMIQLLR
D10023aa SGDGGGLQEQ AEAQGEMEDE AYDEAALGSF VPIEKLQVNG ITMADVKKLR
X64270aa SGDGGGLQEQ AEAQGEMEDE AYDEAALGSF VPIEKLQVNG ITMADVKKLR
U22441aa HRNQKSMQDQ ....ND EIEDVQHGPF .PVEQLQASG IAALDVKKLK
AtU43652aa .RNQNAVQQQ ....DD ..EETQHGPF .PVEQLQAAG IASVDVKKLR
D14134aa ~~~~MAMQMQ LEANA...DT SVEEESFGPQ .PISRLEQCG INANDVKKLE
NM079844aa ~~~~~~~~~~ MEKLTNVQAQ QEEEEEEGP. LSVTKLIGGS ITAKDIKLLQ
1107sid2 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
1107sid6 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
1107sid4 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
ac002387pep ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~MISFGRRKS PAIEETSLAT
AF029669aa MRGKTFRFEM QRDLVSFPLS PAVRVKLVSA GFQTAEELLE VKPSELSKEV
U84138aa ~~~~~~~~~~ ~~~~~~MGSK KLKRVGLSQE LCDRLSRHQI LTCQDFLCLS
101
150
AF034955aa GRKIKTVADL AAADLEEVAQ KCGL YK LV ALRRVLLAQ S FPL GA
AF034956aa SHRIKTVVDL VSADLEEVAQ KCGL YK... ....
D10023aa ESGLHTAEAV AYAPRKDLLE KGI EAKAD KLLNEAARLV PM FVT A F
X64270aa ESGLHTAEAV AYAPRKDLLE KGI EAKAD KLLNEAARLV PM FVT A F
U22441aa DAGLCTVESV VYAPRKELLQ KGI EAKVD KIIEAASKLV PL FTS SQ
AtU43652aa DAGLCTVEGV AYTPRKDLLQ KGI DAKVD KIVEAASKLV PL FTS SQ
D14134aa EAGFHTVEAV AYAPKKELIN KGI EAKAD KILAEAAKLV PM FTT T F
NM079844aa QASLHTVESV ANATKKQLMA PGLGGGKVE QIITEA KLV PL FLS RTF
1107sid2 ~~~~~~~~~~ ~~~~~~~~~~ GDQ .G . ...NGP QK
1107sid6 ~~~~~~~~~~ ~~~~~~~~~~ GDQ .G . ...NGP QK
1107sid4 ~~~~~~~~~~ ~~~~~~~~~~ GDQ .G . ...NGP QK
ac002387pep SVMEAWRLPL SPSIRGKLIS AGYTCLS I. ...ASVSSSD R K
AF029669aa GISKAEALET LQIIRRECLT NKPRYAG S. ...ESHKKC. .....T L
U84138aa PLELMKVTGL SYRGVHELLC VSR CAP . .....M TA GIK R A F
151
200
AF034955aa E LKT TAI I S DK DA YTG V G S C C
AF034956aa ....X DK DA YTG V G S C C
D10023aa HMR.RSELIC KN DT G ETGS F EFRT CHT
X64270aa HMR.RSELIC KN DT G ETGS F EFRT CHT
U22441aa H Q.R EIIQ K DK E ETGS Y EFRC CHT C
AtU43652aa H Q.RQEIIQ R DK E ETGS Y EFRS CHT C
D14134aa HQR.RSEIIQ K DK Q ETGS F EFRT CHT
NM079844aa QM.RADVVQ K DK ETGS F EFRC CHT
1107sid2 . KH N H K
1107sid6 . KH N H K
1107sid4 . KH N H K
ac002387pep LH . E LPR SCS DN S R
AF029669aa LEQ. H GF I FCSA D PLMK T C A C
U84138aa ..... PAF TLSA D A H A GS L T P C FC M


201
250
AF034955aa
AF034956aa
D10023aa
X64270aa
U22441aa
AtU43652aa
D14134aa
NM079844aa
1107sid2
1107sid6
1107sid4
ac002387pep
AF029669aa
U84138aa
251
300
AF034955aa
AF034956aa
D10023aa
X64270aa
U22441aa
AtU43652aa
D14134aa
NM079844aa
1107sid2
1107sid6
1107sid4
ac002387pep
AF029669aa
U84138aa
301
350
AF034955aa
AF034956aa
D10023aa
X64270aa
U22441aa
AtU43652aa
D14134aa
NM079844aa
1107sid2
1107sid6
1107sid4
ac002387pep
AF029669aa
U84138aa


351
400
AF034955aa
AF034956aa
D10023aa
X64270aa
U22441aa
AtU43652aa
D14134aa
NM079844aa
1107sid2
1107sid6
1107sid4
ac002387pep
AF029669aa
U84138aa
401
450
451
472



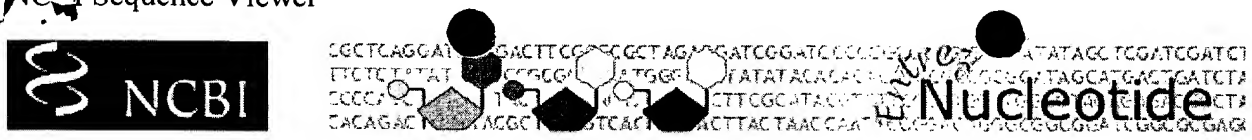
☐ **1:** AF034955. Mus musculus Rad5... [gi:2920579]

Related Sequences, OMIM, Protein, PubMed, Taxonomy, UniSTS, LinkOut

```
LOCUS       AF034955                1699 bp    mRNA    linear   ROD 29-APR-1998
DEFINITION  Mus musculus Rad51d mRNA, complete cds.
ACCESSION   AF034955
VERSION     AF034955.1  GI:2920579
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1699)
  AUTHORS   Pittman,D.L., Weinberg,L.R. and Schimenti,J.C.
  TITLE     Identification, characterization, and genetic mapping of Rad51d, a
            new mouse and human RAD51/RecA-related gene
  JOURNAL   Genomics 49 (1), 103-111 (1998)
  MEDLINE   98234549
REFERENCE   2  (bases 1 to 1699)
  AUTHORS   Pittman,D.L., Weinberg,L.R. and Schimenti,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-1997) The Jackson Laboratory, 600 Main Street,
            Bar Harbor, ME 04609, USA
FEATURES             Location/Qualifiers
     source          1..1699
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     gene            1..1699
                     /gene="Rad51d"
     CDS             196..1185
                     /gene="Rad51d"
                     /codon_start=1
                     /product="RAD51D"
                     /protein_id="AAC40093.1"
                     /db_xref="GI:2920580"
                     /translation="MGMLRAGLCPGLTEETVQLLRGRKIKTVADLAAADLEEVAQKCG
                     LSYKALVALRRVLLAQFSAFPLNGADLYEELKTSTAILSTGIGSLDKLLDAGLYTGEV
                     TEIVGGPGSGKTQVCLCVAANVAHSLQQNVLYVDSNGGMTASRLLQLLQARTQDEEKQ
                     ASALQRIQVVRSFDIFRMLDMLQDLRGTIAQQEATSSGAVKVVIVDSVTAVVAPLLGG
                     QQREGLALMMQLARELKILARDLGVAVVVTNHLTRDWDGRRFKPALGRSWSFVPSTRI
                     LLDVTEGAGTLGSSQRTVCLTKSPRQPTGLQEMIDIGTLGTEEQSPELPGKQT"
BASE COUNT      391 a    456 c    480 g    372 t
ORIGIN
        1 attcggcacg aggcttgcga gttgggaagg gttagtgtcc ctcacctcga cttgcatcct
       61 cttcccgccc cttcgtcccg cgctggcacg ccaggacctt ttccctaagt agaactgagg
      121 aatgcccaga gtgggggact cggcgagcgc ccaagtgaca gagagccccc agggcatcct
      181 gggtttgtag ggactatggg catgctcagg gcagggctgt gcccgggcct caccgaggag
      241 accgtccagc ttctcagagg ccgaaagata aaaacagtgg cagacctggc agctgctgac
      301 ttggaggaag tagcccagaa gtgtggcttg tcctacaagg ccctcgttgc cctgaggagg
      361 gtgttgctgg cgcagttctc ggctttcccc ttaaatggcg cagatctcta tgaggaactg
      421 aagacttcca cggccatcct gtccaccggc atcggaagcc tggacaaact acttgatgct
      481 ggcctctata ctggggaggt gactgaaatt gtgggtggcc caggtagcgg caaaacccag
```

```
      541 gtgtgtctct gtgtggctgc aaatgtggcc catagcctgc agcagaatgt actgtatgtg
      601 gattccaatg gaggaatgac ggcgtcccgc ctcctccagc tactacaggc tagaacccaa
      661 gatgaggaga aacaggcaag tgctctccag aggatacagg tggtgcgttc atttgacatc
      721 ttccggatgc tagatatgct acaggacctt cgcggcacca tagcccagca ggaagcaact
      781 tcttcaggcg ccgtgaaggt tgtgattgtg gactcggtca ctgcagtggt cgccccactt
      841 ctgggaggtc agcagaggga aggcctggcc ttgatgatgc agctggcccg agagctcaag
      901 atcctggccc gggacctggg tgtggcagtg gtggtgacca accacttgac tcgagattgg
      961 gatggtagaa gattcaaacc tgcccttgga cgctcctgga gctttgtgcc cagtacccgg
     1021 attctcctgg atgtcactga gggggctggg acactcggta gcagccaacg cacagtatgt
     1081 ctgaccaagt ctccccgcca gccaacgggt ctgcaggaga tgatagacat tgggacattg
     1141 gggactgagg agcagagccc agaattacct ggcaagcaga cgtgacactg ttgattggga
     1201 aagggacagc ctcaaggccc ccacagcttt ccttcccggg cagctactgt cccctgccac
     1261 atgggaccta gcagtgcaga ggtctcaggc tcatcagaag acaagcctca gtcccctcag
     1321 cttcattgtg tgagatctca tggagcccct ggccagggga tgctgcagtg agagagacct
     1381 agaagtccat cagggtctca agttttcttc ctttgtgtct accgaatgtt tccagtgtgg
     1441 gttgaggtca acctagcttg aggcctcact gaagaaacac agtagtacct gcatgaatag
     1501 tgtcctggtc ctgtgtgtta ccattggttc ctaaacccct ccctcacaca gtaattgggc
     1561 ctttgctcac ccattccccc ccccccattt attttggtct gttttcccat ctgcaagatg
     1621 gcagtttctc agtctgcaag ttattcagta gaaacaacct cagaaatgtc aaaaaaaaaa
     1681 aaaaaaaaaa aaaaaaaaa
//
```

Revised: October 24, 2001.

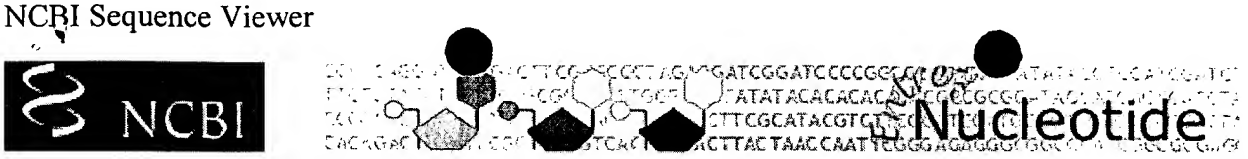


☐ 1: AF034956. Homo sapiens RAD5...[gi:2920581] Related Sequences. OMIM, Protein, PubMed, Taxonomy, LinkOut

```
LOCUS       AF034956                1598 bp    mRNA    linear   PRI 29-APR-1998
DEFINITION  Homo sapiens RAD51D mRNA, complete cds.
ACCESSION   AF034956
VERSION     AF034956.1  GI:2920581
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1598)
  AUTHORS   Pittman,D.L., Weinberg,L.R. and Schimenti,J.C.
  TITLE     Identification, characterization, and genetic mapping of Rad51d, a
            new mouse and human RAD51/RecA-related gene
  JOURNAL   Genomics 49 (1), 103-111 (1998)
  MEDLINE   98234549
REFERENCE   2  (bases 1 to 1598)
  AUTHORS   Pittman,D.L., Weinberg,L.R. and Schimenti,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-1997) The Jackson Laboratory, 600 Main Street,
            Bar Harbor, ME 04609, USA
FEATURES             Location/Qualifiers
     source          1..1598
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            1..1598
                     /gene="RAD51D"
     CDS             125..994
                     /gene="RAD51D"
                     /codon_start=1
                     /product="RAD51D"
                     /protein_id="AAC39719.1"
                     /db_xref="GI:2920582"
                     /translation="MGVLRVGLCPGLTEEMIQLLRSHRIKTVVDLVSADLEEVAQKCG
                     LSYKXLDKLLDAGLYTGEVTEIVGGPGSGKTQVCLCMAANVAHGLQQNVLYVDSNGGL
                     TASRLLQLLQAKTQDEEEQAEALRRIQVVHAFDIFQMLDVLQELRGTVAQQVTGSSGT
                     VKVVVVDSVTAVVSPLLGGQQREGLALMMQLARELKTLARDLGMAVVVTNHITRDRDS
                     GRLKPALGRSWSFVPSTRILLDTIEGAGASGGRRMACLAKSSRQPTGFQEMVDIGTWG
                     TSEQSATLQGDQT"
BASE COUNT      364 a    402 c    455 g    375 t      2 others
ORIGIN
        1 attcggcacg agcgcgcctg tgtcctctct aggaaggggt aggggagggg cgtctggaga
       61 ggaccccccg cgaatgccca cgtgacgtgc agtccccctg gggctgttcc ggcctgcggg
      121 gaacatgggc gtgctcaggg tcggactgtg ccctggcctt accgaggaga tgatccagct
      181 tctcaggagc cacaggatca agacagtggt ggacctggtt tctgcagacc tggaagaggt
      241 agctcagaaa tgtggcttgt cttacaagtn ncttgataaa ctgcttgatg ctggtctcta
      301 tactggagaa gtgactgaaa ttgtaggagg cccaggtagc ggcaaaactc aggtatgtct
      361 ctgtatggca gcaaatgtgg cccatggcct gcagcaaaac gtcctatatg tagattccaa
      421 tggagggctg acagcttccc gcctcctcca gctgcttcag gctaaaaccc aggatgagga
      481 ggaacaggca gaagctctcc ggaggatcca ggtggtgcat gcatttgaca tcttccagat
      541 gctggatgtg ctgcaggagc tccgaggcac tgtggcccag caggtgactg gttcttcagg
```

```
      601 aactgtgaag gtggtggctg tggactcggt cactgcggtg gtttccccac ttctgggagg
      661 tcagcagagg gaaggcttgg ccttgatgat gcagctggcc cgagagctga agaccctggc
      721 ccgggacctt ggcatggcag tggtggtgac caaccacata actcgagaca gggacagcgg
      781 gaggctcaaa cctgccctcg gacgctcctg gagctttgtg cccagcactc ggattctcct
      841 ggacaccatc gagggagcag gagcatcagg cggccggcgc atggcgtgtc tggccaaatc
      901 ttcccgacag ccaacaggtt tccaggagat ggtagacatt gggacctggg ggacctcaga
      961 gcagagtgcc acattacagg gtgatcagac atgacctgtg ctgttgtttg ggaaacaggg
     1021 aagcattggg gacccctccc aacttttctt cccagtaacg cctgctgttt actgccacct
     1081 ggcactggtg actacagacg ttctcaggct ggccagaaga gacatcttgg gttccttggc
     1141 ctcactctct gtaagcatat aaaccacagg cgaaagagga tgctgcattg cgaggaccca
     1201 gaaattcata ctggtgccac gtttccttcc cttatttcta acgtgtatgt ttctggtgga
     1261 aaccaagttc accctggctg ggagcatctc tgatgaggca tgctggcgac tggatggata
     1321 atcctgtgca tcaccattgt gtcctgtgct ccctcctagc gcagtggcca agccgggaaa
     1381 gcctctaact tgcctttgct gctgctgcct ttttttttctt ttgtctctgc ctttccattt
     1441 gttagatggg ggcccactct tccttagctc tgtctctgag ttactgggtg gaaataagct
     1501 tataaatgaa atactcttct tcatctctgt tttgctctta aaaatataaa aaggcaattc
     1561 cccgaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaa
//
```

Revised: October 24, 2001.



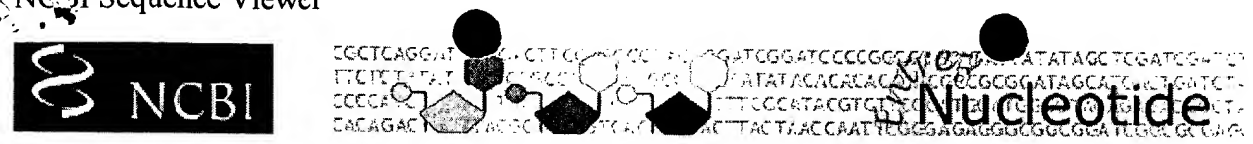
☐ **1:** D10023. S.cerevisiae Rad5...[gi:218468] Related Sequences, Protein, PubMed, Taxonomy

```
LOCUS       YSCRAD51                3724 bp    DNA     linear   PLN 02-FEB-1999
DEFINITION  S.cerevisiae Rad51 protein gene.
ACCESSION   D10023
VERSION     D10023.1  GI:218468
KEYWORDS    Rad51 protein.
SOURCE      Saccharomyces cerevisiae DNA.
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 3724)
  AUTHORS   Shinohara,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-NOV-1991) Akira Shinohara, Faculty of Science, Osaka
            University, Department of Biology; Toyonaka, Osaka 560, Japan
            (E-mail:c62528@center.osaka-u.ac.jp, Tel:06-844-1151(ex.4305),
            Fax:06-841-2449)
REFERENCE   2  (bases 1 to 3724)
  AUTHORS   Shinohara,A., Ogawa,H. and Ogawa,T.
  TITLE     Rad51 protein involved in repair and recombination in S. cerevisiae
            is a RecA-like protein
  JOURNAL   Cell 69 (3), 457-470 (1992)
  MEDLINE   92257587
  REMARK    Erratum:[[published erratum appears in Cell 1992 Oct
            2;71(1):following 180]]
COMMENT     Submitted (29-NOV-1991) to DDBJ by:
            Akira Shinohara
            Department of Biology
            Faculty of Science
            Osaka University
            Toyonaka, Osaka 560
            Japan
            Phone:  06-844-1151 x4305
            Fax:    06-841-244.
FEATURES             Location/Qualifiers
     source          1..3724
                     /organism="Saccharomyces cerevisiae"
                     /db_xref="taxon:4932"
     misc_signal     187..204
                     /note="Box B"
     repeat_region   350..370
                     /rpt_type=direct
                     /rpt_unit=350..359
                     /rpt_unit=361..370
     misc_signal     506..542
                     /note="Box A"
     TATA_signal     569..572
     gene            645..1847
                     /gene="RAD51"
     CDS             645..1847
                     /gene="RAD51"
```

```
                     /codon_start=1
                     /product="Rad51 protein"
                     /protein_id="BAA00913.1"
                     /db_xref="GI:218469"
                     /translation="MSQVQEQHISESQLQYGNGSLMSTVPADLSQSVVDGNGNGSSED
                     IEATNGSGDGGGLQEQAEAQGEMEDEAYDEAALGSFVPIEKLQVNGITMADVKKLRES
                     GLHTAEAVAYAPRKDLLEIKGISEAKADKLLNEAARLVPMGFVTAADFHMRRSELICL
                     TTGSKNLDTLLGGGVETGSITELFGEFRTGKSQLCHTLAVTCQIPLDIGGGEGKCLYI
                     DTEGTFRPVRLVSIAQRFGLDPDDALNNVAYARAYNADHQLRLLDAAAQMMSESRFSL
                     IVVDSVMALYRTDFSGRGELSARQMHLAKFMRALQRLADQFGVAVVVTNQVVAQVDGG
                     MAFNPDPKKPIGGNIMAHSSTTRLGFKKGKGCQRLCKVVDSPCLPEAECVFAIYEDGV
                     GDPREEDE"
     terminator      1927..1932
     CDS             3030..>3724
                     /note="unknown gene product"
                     /codon_start=1
                     /protein_id="BAA20966.1"
                     /db_xref="GI:4433385"
                     /translation="MWPNFTLARKTRRACYQDGKPLKCTGEVVKKTQDIYSNFQYAQY
                     ILRVGLDTEKLHELVKELEDESNSFSVDSLKEYLVNDAKVILKRLSAVGYPDAQYLLG
                     DAYSSGVFGKIKNRRAFLLFSAAAKRMHIESVYRTAICYECGLGVTRNAPKAVNFLTF
                     AATKNHPAAMYKLGVYSYHGLMGLPDDILTKMDGYRWLRRATSMASSFVCGAPFELAN
                     IYMTGYKDLIISDP"
BASE COUNT     1021 a    757 c    806 g   1140 t
ORIGIN      Chromosome V.
        1 ggatccgaca ttttttttt atgctttatt cactgttcaa tattttcacc acaattcgca
       61 agaaacgcac tctacttcga aactacggtt caaacttact tagcagcttc ccgatttaat
      121 tggcctttct actatgccat aaactctttc ttcctctctt ttcatcgccc ctgcatttgc
      181 actttttttgc caccggcagt gccatccggt cacatgacta caccacgtta atagcgatct
      241 ggcttatcat tgtcacagag taaattaaaa tggacggtaa atgttggaaa tgcaccacta
      301 ccgttcttca accaatctag tttagctatc ctgcaacagg tggccttctt gagcattccc
      361 tgagcattcc aaccggttgt atcagtgttt tatcaccgtc tcaccatatc ccacgactag
      421 gccacacttc gttaccctat gctacgcgtc atttccgcta tttctgtcct ggtttgttta
      481 cagtacgcgt ggtgggacca taaaggggaa tagtggggac tggagaaaaa attttctcag
      541 ttacttcttc tatcttccgt agtttccata tactagtagt tgagtgtagc gacaaagagc
      601 agacgtagtt atttgttaaa ggcctactaa tttgttatcg tcatatgtct caagttcaag
      661 aacaacatat atcagagtca cagcttcagt acgggaacgg ttcgttgatg tccactgtac
      721 cagcagacct ttcacagtca gtcgttgatg gaaacggcaa cggtagcagc gaagatattg
      781 aggccaccaa cggctccggc gatggtggcg gattgcagga gcaagcggaa gcgcaaggtg
      841 aaatggagga tgaagcatac gatgaagctg ccttaggttc gtttgtgcca atagaaaaac
      901 tgcaagtgaa cgggattact atggcggatg tgaaaaaact aagggagagt gggcttcaca
      961 ctgctgaagc ggtagcatat gctcccagaa aggatttatt ggaaatcaaa ggtatatcgg
     1021 aagctaaggc agataagttg ctaaacgaag cggcaaggct agtgcctatg ggatttgtca
     1081 cggctgctga ttttcatatg agaagatcgg agctgatttg tttgacaacg ggttctaaga
     1141 atttggacac tcttttgggt ggtggtgtgg aaactggttc tattactgag cttttcggtg
     1201 aattcaggac aggtaagtcc cagctatgtc acactttggc cgtgacatgc caaattccat
     1261 tggatattgg tggcggtgaa ggtaagtgtt tgtatatcga taccgaaggt actttcaggc
     1321 cggtaagatt ggtatccata gctcagcggt tcggattaga cccggatgat gctttgaaca
     1381 acgttgcgta tgcaagagcc tataacgccg atcatcagtt aagacttctg gatgctgctg
     1441 cccaaatgat gagcgagtct cggttttcct tgattgtggt cgattctgtt atggctctat
     1501 accgtacgga tttttctggt cgtggtgaac taagcgcaag gcaaatgcat ttagccaaat
     1561 ttatgcgtgc tttgcaaagg ctggccgacc aatttggtgt tgcagtcgtc gttactaacc
     1621 aagtggtcgc ccaagttgat ggtggtatgg cttttaatcc agatccaaag aagcctatcg
     1681 gtggtaatat tatggcacat tcttccacca cgcgattagg tttcaaaaag ggtaagggat
     1741 gtcaaagatt atgcaaagtt gttgactcac cttgcttacc agaggctgaa tgtgtgttcg
     1801 cgatctatga agatggtgtt ggtgacccca gagaagaaga cgagtaggta tttggtctct
     1861 tgtctctatt tatttacaca ggtttacttt caattctcct ctttttctta ggttgcgttc
     1921 cgtacatttt tatcttcatt tccatccact gtcttagatt tttgcatata ttttgtcata
     1981 tacctcgcaa ccctactgcg gtcttaacct tttttttcag ttcttttaaa taactttcgt
     2041 attgtctgtc acccatgaaa tatatgtatt tttctactct tcttcccgat gactacttcc
     2101 tcctgcaggg tccgcgcggc tttatccttt taggggagtg aagagaaaaa ttttctgata
     2161 tgtcgccatc ctcttataaa accaacgtaa aataccatac acttctttat ggcaatgaat
     2221 ttccttgtct tcgtttcgac gtttttaaat tgaataaaaa attgtacttg gaaaaaagaa
```

```
     2281 tagaacaata cttttgctaa cagcaagaga agtacggttg acattacgag tatccctttg
     2341 ttaaaatcag gcagtaccca taatgtccat gactatttgc tcaaatactc ctggtgcata
     2401 ccctgaaatt ggagcttata atgagggtga taacagttgg aatcaagcgg attctcccct
     2461 gattcctctc tgatattgaa taaaccagaa gttcgtcaat actggtcatc cgtttcatct
     2521 catatttcac gctcaggtga tgtgtttaca aatgataaag aaaaaatttc atcctctatt
     2581 ggtgaggatg cgatggatat cgacgcttcg ccctcattga ttgaaaaaat aactcttttc
     2641 ccacaagaaa gattcttcct gagcaggacg agtttgagaa cgacgttgag gacgatgctt
     2701 cttcctcact gaaggaaaaa tcacaaggtt cttgtgaaat tgaaattgct tctgaaattt
     2761 catccgaaat tttaaatggc acatcggcag acggtaattc cgagtttcac gacttcgctg
     2821 agccccctcc ctctcagaat gaatctgtcg ctctgtcttt tagtcaatcg aatgatttgg
     2881 acttttcgaa taatccaagc ggatcaggct cttcaacgat atcaatagaa gtacaagttc
     2941 catttcactt cctagacatg tgtcactaga ttttaacgtt tacaatagtc tttgcctcac
     3001 aaatgaggtg actgcatcag aatcacataa tgtggccaaa tttcaccttg gcaaggaaaa
     3061 caagaagagc ttgctaccaa gatggaaaac cattgaaatg taccggggag gtagttaaga
     3121 aaactcaaga catatattct aattttcagt atgcacaata cattttaagg gtagggttgg
     3181 ataccgaaaa actacacgaa cttgttaagg aactggagga tgaaagcaat agcttctctg
     3241 tagattcttt aaaggaatat ttagtgaacg atgccaaagt tattttgaag agactaagtg
     3301 ccgttggata ccctgatgca cagtaccttc taggtgacgc ctactcctct ggagttttcg
     3361 gtaaaataaa gaatagaaga gcatttcttt tattctccgc tgcagccaaa agaatgcaca
     3421 tcgaaagtgt ttacagaact gccatttgct atgagtgcgg cttgggtgta accagaaatg
     3481 caccgaaggc ggttaacttt ttgacttttg ctgcaactaa gaaccatcct gcagcaatgt
     3541 acaaattagg agtatattcg tatcatggtt tgatgggtct tccagatgat attctaacca
     3601 aaatggatgg ttatagatgg ttgcgaaggg ctacatctat ggctagcagc tttgtttgtg
     3661 gtgctccttt tgaattagcc aatatttata tgacgggata taaagatctt atcatttcgg
     3721 atcc
//
```

Revised: October 24, 2001.



1: X64270. S.cerevisiae RAD5...[gi:4274]

Related Sequences, Protein, PubMed, Taxonomy

```
LOCUS       SCRADDNA                2173 bp    DNA     linear   PLN 03-DEC-1993
DEFINITION  S.cerevisiae RAD51 gene.
ACCESSION   X64270 S38937
VERSION     X64270.1  GI:4274
KEYWORDS    DNA repair; RAD51 gene; recombination and repair.
SOURCE      baker's yeast.
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 2173)
  AUTHORS   Fabre,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JAN-1992) F. Fabre, Institut Curie, Section de
            Biologie, Centre Universitaire, Batiment 110, 91405 Orsay, FRANCE
REFERENCE   2  (bases 1 to 2173)
  AUTHORS   Aboussekhra,A., Chanet,R., Adjiri,A. and Fabre,F.
  TITLE     Semidominant suppressors of Srs2 helicase mutations of
            Saccharomyces cerevisiae map in the RAD51 gene, whose sequence
            predicts a protein with similarities to procaryotic RecA proteins
  JOURNAL   Mol. Cell. Biol. 12 (7), 3224-3234 (1992)
  MEDLINE   92318940
FEATURES             Location/Qualifiers
     source          1..2173
                     /organism="Saccharomyces cerevisiae"
                     /strain="S288C"
                     /sub_strain="GRF88"
                     /db_xref="taxon:4932"
                     /chromosome="VR"
                     /map="80cM"
     misc_signal     444..449
                     /note="DNA synthesis control sequence"
     misc_signal     485..490
                     /note="DNA synthesis control sequence"
     gene            645..1847
                     /gene="RAD51"
     CDS             645..1847
                     /gene="RAD51"
                     /function="repair of DNA double  strand breaks"
                     /note="similarities to procaryotic  RecA"
                     /codon_start=1
                     /protein_id="CAA45563.1"
                     /db_xref="GI:4275"
                     /db_xref="SWISS-PROT:P25454"
                     /translation="MSQVQEQHISESQLQYGNGSLMSTVPADLSQSVVDGNGNGSSED
                     IEATNGSGDGGGLQEQAEAQGEMEDEAYDEAALGSFVPIEKLQVNGITMADVKKLRES
                     GLHTAEAVAYAPRKDLLEIKGISEAKADKLLNEAARLVPMGFVTAADFHMRRSELICL
                     TTGSKNLDTLLGGGVETGSITELFGEFRTGKSQLCHTLAVTCQIPLDIGGGEGKCLYI
                     DTEGTFRPVRLVSIAQRFGLDPDDALNNVAYARAYNADHQLRLLDAAAQMMSESRFSL
                     IVVDSVMALYRTDFSGRGELSARQMHLAKFMRALQRLADQFGVAVVVTNQVVAQVDGG
                     MAFNPDPKKPIGGNIMAHSSTTRLGFKKGKGCQRLCKVVDSPCLPEAECVFAIYEDGV
```

GDPREEDE

terminator 2063..2067

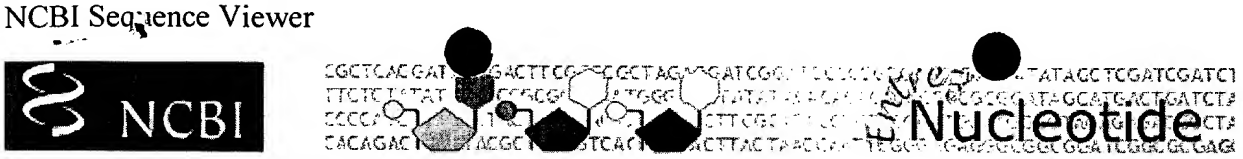
BASE COUNT 541 a 461 c 496 g 675 t

ORIGIN

```
        1 ggatccgaca ttttttttt atgctttatt cactgttcaa tattttcacc acaattcgca
       61 agaaacgcac tctacttcga aactacggtt caaacttact tagcagcttc ccgatttaat
      121 tggcctttct actatgccat aaactctttc ttcctctctt ttcatcgccc ctgcatttgc
      181 actttttgc caccggcagt gccatccggt cacatgacta caccacgtta atagcgatct
      241 ggcttatcat tgtcacagag taaattaaaa tggacggtaa atgttggaaa tgcaccacta
      301 ccgttcttca accaatctag tttagctatc ctgcaacagg tggccttctt gagcattccc
      361 tgagcattcc aaccggttgt atcagtgttt tatcaccgtc tcaccatatc ccacgactag
      421 gccacacttc gttaccctat gctacgcgtc atttccgcta tttctgtcct ggtttgttta
      481 cagtacgcgt ggtgggacca taaaggggaa tagtggggac tggagaaaaa attttctcag
      541 ttacttcttc tatcttccgt agtttccata tactagtagt tgagtgtagc gacaaagagc
      601 agacgtagtt atttgttaaa ggcctactaa tttgttatcg tcatatgtct caagttcaag
      661 aacaacatat atcagagtca cagcttcagt acgggaacgg ttcgttgatg tccactgtac
      721 cagcagacct ttcacagtca gtcgttgatg gaaacggcaa cggtagcagc gaagatattg
      781 aggccaccaa cggctccggc gatggtggcg gattgcagga gcaagcggaa gcgcaaggtg
      841 aaatggagga tgaagcatac gatgaagctg ccttaggttc gtttgtgcca atagaaaaac
      901 tgcaagtgaa cgggattact atggcggatg tgaaaaaact aagggagagt gggcttcaca
      961 ctgctgaagc ggtagcatat gctcccagaa aggatttatt ggaaatcaaa ggtatatcgg
     1021 aagctaaggc agataagttg ctaaacgaag cggcaaggct agtgcctatg ggatttgtca
     1081 cggctgctga ttttcatatg agaagatcgg agctgatttg tttgacaacg ggttctaaga
     1141 atttggacac tcttttgggt ggtggtgtgg aaactggttc tattactgag cttttcggtg
     1201 aattcaggac aggtaagtcc cagctatgtc acactttggc cgtgacatgc caaattccat
     1261 tggatattgg tggcggtgaa ggtaagtgtt tgtatatcga taccgaaggt actttcaggc
     1321 cggtaagatt ggtatccata gctcagcggt tcggattaga cccggatgat gctttgaaca
     1381 acgttgcgta tgcaagagcc tataacgccg atcatcagtt aagacttctg gatgctgctg
     1441 cccaaatgat gagcgagtct cggttttcct tgattgtggt cgattctgtt atggctctat
     1501 accgtacgga tttttctggt cgtggtgaac taagcgcaag gcaaatgcat ttagccaaat
     1561 ttatgcgtgc tttgcaaagg ctggccgacc aatttggtgt tgcagtcgtc gttactaacc
     1621 aagtggtcgc ccaagttgat ggtggtatgg cttttaatcc agatccaaag aagcctatcg
     1681 gtggtaatat tatggcacat tcttccacca cgcgattagg tttcaaaaag ggtaagggat
     1741 gtcaaagatt atgcaaagtt gttgactcac cttgcttacc agaggctgaa tgtgtgttcg
     1801 cgatctatga agatggtgtt ggtgacccca gagaagaaga cgagtaggta tttggtctct
     1861 tgtctctatt tatttacaca ggtttacttt caattctcct ctttttctta ggttgcgttc
     1921 cgtacatttt tatcttcatt tccatccact gtcttagatt tttgcatata ttttgtcata
     1981 tacctcgcaa ccctactgcg gtcttaacct tttttttcag ttcttttaaa taactttcgt
     2041 attgtctgtc acccatgaaa tatatgtatt tttctactct tcttcccgat gactacttcc
     2101 tcctgcaggg tccgcgcggc tttatccttt taggggagtg aagagaaaaa ttttctgata
     2161 tgtcgccatc ctc
```

//

Revised: October 24, 2001.



1: U22441. Lycopersicon escu...[gi:1143809] Protein, PubMed, Taxonomy, LinkOut

```
LOCUS       LEU22441                1241 bp    mRNA    linear   PLN 18-JUN-1998
DEFINITION  Lycopersicon esculentum LeRAD51 (RAD51) mRNA, complete cds.
ACCESSION   U22441
VERSION     U22441.1  GI:1143809
KEYWORDS    .
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon.
REFERENCE   1  (bases 1 to 1241)
  AUTHORS   Yeager Stassen,N., Logsdon,J.M. Jr., Vora,G.J., Offenberg,H.H.,
            Palmer,J.D. and Zolan,M.E.
  TITLE     Isolation and characterization of rad51 orthologs from Coprinus
            cinereus and Lycopersicon esculentum, and phylogenetic analysis of
            eukaryotic recA homologs
  JOURNAL   Curr. Genet. 31 (2), 144-157 (1997)
  MEDLINE   97174112
REFERENCE   2  (bases 1 to 1241)
  AUTHORS   Yeager Stassen,N., Logsdon,J.M. Jr., Vora,G.J., Offenberg,H.H.,
            Palmer,J.D. and Zolan,M.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-1995) Biology, Indiana University, Bloomington,
            IN 47405, USA
FEATURES             Location/Qualifiers
     source          1..1241
                     /organism="Lycopersicon esculentum"
                     /db_xref="taxon:4081"
                     /tissue_type="anther"
     gene            1..1241
                     /gene="RAD51"
     CDS             98..1126
                     /gene="RAD51"
                     /note="similar to human RAD51 protein: PIR Accession
                     Number S37673 and to RecA-like protein, encoded by Genbank
                     Accession Number Z24756"
                     /codon_start=1
                     /product="LeRAD51"
                     /protein_id="AAC23700.1"
                     /db_xref="GI:1143810"
                     /translation="MEQQHRNQKSMQDQNDEIEDVQHGPFPVEQLQASGIAALDVKKL
                     KDAGLCTVESVVYAPRKELLQIKGISEAKVDKIIEAASKLVPLGFTSASQLHAQRLEI
                     IQITSGSKELDKILEGGIETGSITEIYGEFRCGKTQLCHTLCVTCQLPLDQGGGEGKA
                     MYIDAEGTFRPQRLLQIADRYGLNGPDVLENVAYARAYNTDHQSRLLLEAASMMVETR
                     FALMIVDSATALYRTDFSGRGELSARQMHLAKFLRSLQKLADEFGVAVVITNQVVAQV
                     DGSAVFAGPQIKPIGGNIMAHASTTRLALRKGRAEERICKVVSSPCLAEAEARFQISV
                     EGVTDVKD"
     polyA_site      1241
                     /gene="RAD51"
```

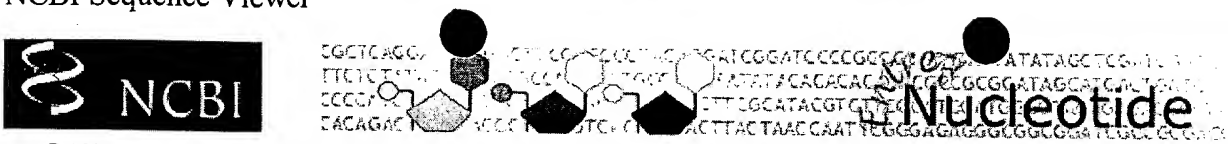
/note="24 A nucleotides"

BASE COUNT 361 a 227 c 302 g 351 t

ORIGIN

```
        1 gaattcggca cgagattttt ggcggctcct cacgcctttg aaatatttcc ttttgcctcc
       61 attcttccat tctcagtgag catcaagaaa attagcaatg gagcagcagc acaggaatca
      121 gaagtcgatg caagaccaaa atgatgaaat cgaggatgtt caacacggcc cttttccagt
      181 tgaacaactt caggcatcag ggattgcagc tctagatgta aaaaaactca aggatgctgg
      241 tctatgtaca gttgaatctg ttgtttatgc tccaagaaag gaacttctgc agataaaagg
      301 aattagtgaa gctaaagttg acaagattat tgaggcagct tcaaaattag tgcctttggg
      361 attcactagt gccagccaac tccatgcaca gaggcttgaa atcatacaga taacttctgg
      421 atcgaaagaa cttgacaaga tattagaagg aggaatcgaa actggatcta ttactgaaat
      481 ttacggagag ttccgatgtg gaaagactca gctgtgtcac acactatgcg tgacttgtca
      541 acttccatta gatcagggag gtggtgaagg gaaagcaatg tacattgatg ctgagggtac
      601 tttcagacca caaagacttt tacaaattgc agacaggtat ggattgaatg gtcctgatgt
      661 cctggagaat gtagcctatg ctcgagctta taataccgat catcaatcaa gacttttgct
      721 tgaggcagcc tcaatgatgg tggagaccag gtttgctctc atgattgtgg acagtgctac
      781 tgccctttat agaactgact tctctgggag aggagagttg tctgccaggc agatgcatct
      841 tgcaaagttt ctgagaagcc ttcagaagtt agcagatgag tttggtgttg ctgttgttat
      901 tacgaaccaa gttgttgctc aagtggatgg ttctgctgta tttgctgggc ctcaaataaa
      961 acccattggt ggcaacatca tggcacatgc ttctacgacg agactagctc tgaggaaggg
     1021 tagggccgag gagcggattt gtaaagtagt cagttcgcca tgcttagctg aagcagaagc
     1081 aagatttcaa atttctgttg aaggagtcac tgatgtaaag gactaaatgt gtgatcagca
     1141 cattgtttac tactagtact atttttgtt tcctacttgg tgtacgattt tgtcatcgtt
     1201 ttgaaggtta gttaaccata aaaagaagta tgctatatgg a
//
```

Revised: October 24, 2001.



1: U43652. Arabidopsis thali...[gi:1706948] Related Sequences, Protein, PubMed, Taxonomy, LinkOut

```
LOCUS       ATU43652                4868 bp    DNA     linear   PLN 04-DEC-1996
DEFINITION  Arabidopsis thaliana RAD51 homolog (AtRad51) gene, complete cds,
            and tRNA-Cys gene, complete sequence.
ACCESSION   U43652
VERSION     U43652.1  GI:1706948
KEYWORDS    .
SOURCE      thale cress strain=Landsberg erecta.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 4868)
  AUTHORS   Urban,C., Smith,K.N. and Beier,H.
  TITLE     Nucleotide sequences of nuclear tRNA(Cys) genes from Nicotiana and
            Arabidopsis and expression in HeLa cell extract
  JOURNAL   Plant Mol. Biol. 32 (3), 549-552 (1996)
  MEDLINE   97134945
REFERENCE   2  (bases 1 to 4868)
  AUTHORS   Smith,K.N., Shinohara,A. and Signer,E.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-DEC-1995) Kathleen N. Smith, Department of Biology,
            MIT, Cambridge, MA 02139, USA
FEATURES             Location/Qualifiers
     source          1..4868
                     /organism="Arabidopsis thaliana"
                     /strain="Landsberg erecta"
                     /db_xref="taxon:3702"
                     /chromosome="5"
                     /map="linked TSL and mi138"
     tRNA            197..268
                     /note="designated pAtC1"
                     /product="tRNA-Cys"
     tRNA            603..674
                     /note="designated pAtC2"
                     /product="tRNA-Cys"
     tRNA            1016..1087
                     /note="designated pAtC3"
                     /pseudo
     mRNA            join(1708..1853,2258..2402,2513..2623,2698..2789,
                     2921..3015,3101..3214,3303..3432,3520..3641,3737..3958)
                     /gene="AtRad51"
                     /note="RAD51 homolog"
                     /product="AtRAD51"
     gene            1708..3958
                     /gene="AtRad51"
     CDS             join(1758..1853,2258..2402,2513..2623,2698..2789,
                     2921..3015,3101..3214,3303..3432,3520..3641,3737..3860)
                     /gene="AtRad51"
                     /note="RAD51 homolog; the translational start site has not
                     been determined, there are two in-frame Methionine
```

residues in the predicted protein sequence"
/codon_start=1
/product="AtRAD51"
/protein_id="AAC49555.1"
/db_xref="GI:1706949"
/translation="MTTMEQRRNQNAVQQQDDEETQHGPFPVEQLQAAGIASVDVKKL
RDAGLCTVEGVAYTPRKDLLQIKGISDAKVDKIVEAASKLVPLGFTSASQLHAQRQEI
IQITSGSRELDKVLEGGIETGSITELYGEFRSGKTQLCHTLCVTCQLPMDQGGGEGKA
MYIDAEGTFRPQRLLQIADRFGLNGADVLENVAYARAYNTDHQSRLLLEAASMMIETR
FALLIVDSATALYRTDFSGRGELSARQMHLAKFLRSLQKLADEFGVAVVITNQVVAQV
DGSALFAGPQFKPIGGNIMAHATTTRLALRKGRAEERICKVISSPCLPEAEARFQIST
EGVTDCKD"

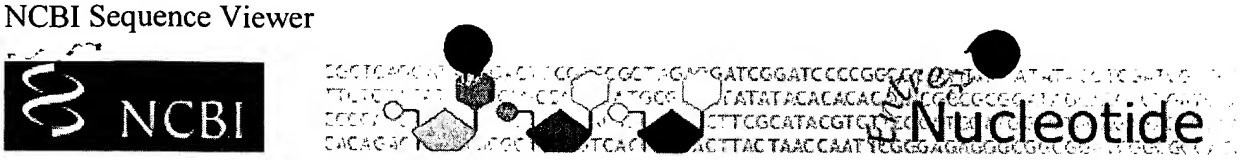
BASE COUNT 1469 a 745 c 879 g 1775 t

ORIGIN

```
       1 tttgtatata attctttagc aagtgaatat gtttttcttt ataatttgaa ggtttaattt
      61 gtttgtgaaa attgtttttg ataatttata tgattactaa ataagtaaac aattgacttg
     121 cttatattag atttcttagc aaaaaaacaa ttaatgaaat aaacaattta ttattttgaa
     181 cttattaaag caataaaggt ccatagctca gtggtagagc aattgactgc agatcaatag
     241 gtcaccggtt cgaacccggt tgggccctat atttttcag tttaccaata atttttaata
     301 agagctttag ttatatattt gattagcatt tagcgtcaag tagttgacta atcttatatt
     361 tcaaggtttt agttagtaat ttttcatctt gaataaaata aattttgaca agtttttgta
     421 tataattctt gagcaactga aaatattttt ttaaaatttc atggtttagt ttgtttgtta
     481 tttgttttat gaaataatta tgtgatttct aaataagtaa ataattgact atattagatt
     541 ttttagcaaa aaaacaattg acgaaataaa taatttacaa tttgaactta ttaaagcaat
     601 aagggtccat agctcagtgg tagagcaatt gactgcagat caataggtca ccggttcgaa
     661 cccggttggg ccctatattt ttcagtttac caatattttt tataagagct ttattaatat
     721 atttgattaa catttagcgt ctactagttg actaatttta tattataagg ttttttctcg
     781 caatttttca tcttgaataa aaacattttg acaagttttt taatataatt cattagcaac
     841 tgaaaatgtt tttttttca tttcatggtt taatttgttt gtgaaaattg tttttgataa
     901 tttatatgat tactaaataa gtaaataact gacttggata tattagattt cttagcaaaa
     961 aataaaataa ttaatagata aataatttat gattttgaac ttattaaagt aataagggtt
    1021 cataactcag tggtagagca attgactgca gatcaatagg tcaccggttt gaacccggtt
    1081 aggccctata ttttttagtt taccaaaaaa atttaaatat catcttgaat aaagaaaatt
    1141 gacaaatttt gtgatatttg taatatttta tttttattat aataagtgat tactacattg
    1201 ttggaattgt ggtggttctc ggcggtcaaa cacctaggta ccatttggtt gacattcaaa
    1261 cacctaggta ccactcggcg gtcaaacacc tattgttttt acaaaacgtt aatttagtgt
    1321 tttaaaaata tataatttta agtaaaaata atttaaaata aaaaaataat tttgagaatc
    1381 gataattcga tcaactttga taaatatcta tcatttataa tttcatgcat ttaactgaaa
    1441 atttaaaatt actatggtac ttaattaata ataaaaatga ggaggatttt gttgttgttt
    1501 ttgagtattt tatagaataa gaatttgggc tttaatagcc tttaaagccc aatatgatca
    1561 aggccgagga aaagctgacc caaacgtaat cgagacttgt tgaagaagcc tttgccctca
    1621 tcgtcgtctt gtataataat tttggttgtg gcgcttcttt caatttgttt tcagtttcgc
    1681 catttccctc cactctcaag ctctcttttg cttctctcgc tttctctggt gacccgaatc
    1741 tgctctgatt gagagaaatg acgacgatgg agcagcgtag aaaccagaat gctgtccaac
    1801 aacaagacga tgaagaaacc cagcacggac ctttccctgt cgaacagctt caggttcaat
    1861 ttctgcagat ttctttcatt tgttgtggaa atttaccttc gcagtctata cgcttcatca
    1921 gatcttgcgc taattgaaat gggttcggct ttaattcact aaaaatttct gctttttttg
    1981 tacatcgcga tgtgaaatag gtgtgaatct gggcgagaat tgagtttcgc tactgtctgt
    2041 taggcgaatt ggttttaggg gttcgttcaa ttcagaaaaa aatcgacact ttttgggggt
    2101 tttatccttt tctggggaac ttctctattg ctgtgactcc agtgtcatat aatctctata
    2161 gttcttctca attcgtgttc ataaaaggga gatgcatcag tgtgttttgt gttatgatta
    2221 tatgctattt catctcttta aatcttcaaa acttcaggca gcaggtattg cttctgttga
    2281 tgtaaagaag cttagggatg ctggtctctg tactgttgaa ggtgttgctt atactccgag
    2341 gaaggatctc ttgcagatta aaggaattag tgatgccaag gttgacaaga ttgtagaagc
    2401 aggtattaca catgtagaaa ctttttgctt tccttctctg ttaaatacat aacacctctt
    2461 tgatactctt agagttaatg tgtgttctta tttgtgtttt cttctgtgat agcttcaaag
    2521 ctagttcctc tggggttcac tagtgcgagc cagctccatg ctcagagaca ggaaattatt
    2581 cagattacct ctggatcacg ggagctcgat aaagttctag aaggttggtg ctggtttctg
    2641 gtgcatctta taacggcatt ttgttattgt gatcttactt tgccgtatgc tcaacaggag
    2701 gtattgaaac tggttccatc acggagttat atggtgagtt ccgctctgga aagactcagc
    2761 tgtgccatac actgtgtgtg acttgtcaag taagaaccct gttcctatac cactcagtta
    2821 ttgttgctta tctagcagaa atcagtgatc ttgtctttct tcttacaatt tccaaacctg
    2881 agaaagatat tcacattaag gttcacgttt gaatgattag cttcccatgg atcaaggagg
```

```
     2941 tggagaggga aaggccatgt acattgatgc tgagggaaca ttcaggccac aaagactctt
     3001 acagatagct gacaggtatg ttattcttgt aacacacatg agatcacaat agcgttcttg
     3061 agacggaatc atctgtaaga aatcaaatga aatgatgcag gtttggatta aatggagctg
     3121 atgtactaga aaacgttgcc tatgcgaggg cgtataatac agatcatcag tcaaggcttt
     3181 tgcttgaagc agcatcaatg atgattgaaa caaggtgtgt tgagtttatt ttggtgttca
     3241 gttccgttat gtctatcatg agagtgctaa cttttaaagt tacgatgacc ctattggcac
     3301 aggtttgctc tcctgattgt cgatagtgct accgctctct acagaacaga tttctctgga
     3361 aggggagagc tttcggctcg acaaatgcat cttgcaaagt tcttgagaag tcttcagaag
     3421 ttagcagatg aggtgaacat tcagtactaa ctgtttcttt tccaatattg ctctgagaca
     3481 ccatcctgaa gtattctctc ttatgtttaa tggctctagt ttggtgtggc tgttgttata
     3541 acaaaccaag tagttgcgca agtagatggt tcagctcttt ttgctggtcc ccaatttaag
     3601 ccgattggtg ggaatatcat ggctcatgcc accacaacaa ggtctgtaat gtttgcaaat
     3661 tcgacaaatc catgtttctt agtgtttttt gttagggttc gttaggaaag tgtgtttgtt
     3721 atgataattg atacaggttg gcgttgagga aaggaagagc agaggagaga atctgtaaag
     3781 tgataagctc gccatgtttg ccagaagcgg aagctcgatt tcaaatatct acagaaggtg
     3841 taacagattg caaggattga tgtttttggc cacaacaatg ccttgcttct ctctgcctta
     3901 tgtgcttttt tttagcaacc ccttcacttg tattgatgtc aaaaaaccgt tttcttttgt
     3961 ttttgttttt aactgttgat catagtgtaa gcgtacaagt aagcgatgat ggagcaaaag
     4021 agctttgtaa gtaaagtttt tcgggtaatg aacctttgt gtttcacaaa cgcacacagc
     4081 ctcgtggttt atacaacact atatattact tagtgaacaa tcaaatatta actaaccatt
     4141 acttaccatt agaatgtctt atggtttatc atgtttttta tatactaata ggcatttaat
     4201 cacatttctt tttatttaac aacgaaaaat ccatattatt acaattgtaa aacataaaag
     4261 catcaccatc tccaatggtt taattagtaa ttgtgattca catattgtga ttctataaac
     4321 tcacaaattg tgattcgata gaagtctaat tattataata aaaattgtca taagaacact
     4381 aaaatagttt atcatgtttt tgtcatacta tctcagataa ttttcatttt ttctgttgtt
     4441 gacatcacca tctccaatat tcaaaacatg atttgtttgt acctttttag tattaacaaa
     4501 acttttgtta gaatgactca tcctctgcat atatataaaa aaatcattag tatttttata
     4561 tcaagtctac aaattatttt catttacttt attatttttt tttaccaaac acatttagtt
     4621 ctatatacta tttattctta ttaactaaag ttactcagaa ttataggtgt aactggaaaa
     4681 ctggataaag agagaaagag agagagtaca taacataagt cctcgaaagt gcagctattc
     4741 cattagcggc cagacagctg ttagtatata ttgcttcgtc acgggctctt ttttttttct
     4801 ttttaattta aagatccctt aaaaaaatct tttttgtttt aaagaagaga gaaattgtgg
     4861 ttttgttg
//
```

Revised: October 24, 2001.



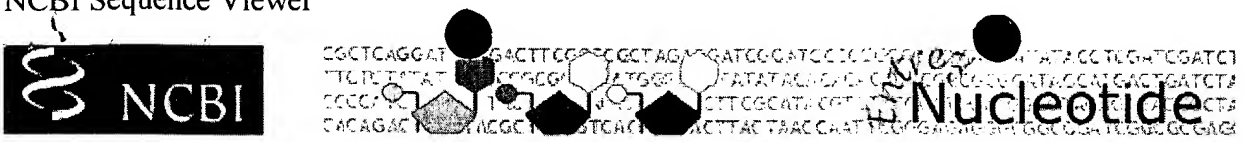
☐ **1:** D14134. Human mRNA for RA... [gi:285976]

ProbeSet, Related Sequences, OMIM, Protein, PubMed, Taxonomy, UniSTS, LinkOut

```
LOCUS       HUMRAD51                2229 bp    mRNA    linear   PRI 03-FEB-1999
DEFINITION  Human mRNA for RAD51, complete cds.
ACCESSION   D14134
VERSION     D14134.1  GI:285976
KEYWORDS    RAD51; histone H2A.
SOURCE      Homo sapiens (sub_species:Caucasian) testis cDNA to mRNA,
            clone_lib:cDNA in pCD8.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2229)
  AUTHORS   Morita,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JAN-1993) Takashi Morita, Research Institute for
            Microbial Deseases, Dept. of Microbial Genetics, Osaka Univ.; 3-1
            Yamadaoka, Suita, Osaka 565, Japan (Tel:06-877-5121(ex.3172),
            Fax:06-876-2678)
REFERENCE   2  (bases 1 to 2229)
  AUTHORS   Yoshimura,Y., Morita,T., Yamamoto,A. and Matsushiro,A.
  TITLE     Cloning and sequence of the human RecA-like gene cDNA
  JOURNAL   Nucleic Acids Res. 21 (7), 1665 (1993)
  MEDLINE   93241950
COMMENT     Submitted (22-JAN-1993) to DDBJ by:
            Takashi Morita
            Department of Microbial Genetics
            Research Institute for Microbial Diseases
            Osaka University
            3-1 Yamadaoka
            Suita, Osaka 565
            Japan
            Phone:  06-875-2913
            Fax:    06-876-2678.
FEATURES             Location/Qualifiers
     source          1..2229
                     /organism="Homo sapiens"
                     /sub_species="Caucasian"
                     /db_xref="taxon:9606"
                     /tissue_type="testis"
                     /clone_lib="cDNA in pCD8"
     gene            233..1252
                     /gene="Rad51"
     CDS             233..1252
                     /gene="Rad51"
                     /codon_start=1
                     /product="RAD51"
                     /protein_id="BAA03189.1"
                     /db_xref="GI:285977"
                     /translation="MAMQMQLEANADTSVEEESFGPQPISRLEQCGINANDVKKLEEA
                     GFHTVEAVAYAPKKELINIKGISEAKADKILAEAAKLVPMGFTTATEFHQRRSEIIQI
```

```
                     TTGSKELLKLLQGGIETGSITEMFGEFRTGKTQICHTLAVTCQLPIDRGGGEGKAMYI
                     DTEGTFRPERLLAVAERYGLSGSDVLDNVAYARAFNTDHQTQLLYQASAMMVESRYAL
                     LIVDSATALYRTDYSGRGELSARQMHLARFLRMLLRLADEFGVAVVITNQVVAQVDGA
                     AMFAADPKKPIGGNIIAHASTTRLYLRKGRGETRICKIYDSPCLPEAEAMFAINADGV
                     GDAKD"
     repeat_region   1862..2173
                     /rpt_unit=1862..1872
     polyA_signal    2208..2213
BASE COUNT      593 a    472 c    602 g    562 t
ORIGIN      Chromosome 15.
        1 ccgcgcgcag cggccagaga ccgagcccta aggagagtgc ggcgcttccc gaggcgtgca
       61 gctgggaact gcaactcatc tgggttgtgc gcagaaggct ggggcaagcg agtagagaag
      121 tggagcgtaa gccaggggcg ttgggggccg tgcgggtcgg gcgcgtgcca cgcccgcggg
      181 gtgaagtcgg agcgcggggc ctgctggaga gaggagcgct gcggaccgag taatggcaat
      241 gcagatgcag cttgaagcaa atgcagatac ttcagtggaa gaagaaagct ttggcccaca
      301 acccatttca cggttagagc agtgtggcat aaatgccaac gatgtgaaga aattggaaga
      361 agctggattc catactgtgg aggctgttgc ctatgcgcca aagaaggagc taataaatat
      421 taagggaatt agtgaagcca aagctgataa aattctggct gaggcagcta aattagttcc
      481 aatgggtttc accactgcaa ctgaattcca ccaaaggcgg tcagagatca tacagattac
      541 tactggctcc aaagagcttg acaaactact tcaaggtgga attgagactg gatctatcac
      601 agaaatgttt ggagaattcc gaactgggaa gacccagatc tgtcatacgc tagctgtcac
      661 ctgccagctt cccattgacc ggggtggagg tgaaggaaag gccatgtaca ttgacactga
      721 gggtaccttt aggccagaac ggctgctggc agtggctgag aggtatggtc tctctggcag
      781 tgatgtcctg gataatgtag catatgctcg agcgttcaac acagaccacc agacccagct
      841 cctttatcaa gcatcagcca tgatggtaga atctaggtat gcactgctta ttgtagacag
      901 tgccaccgcc ctttacagaa cagactactc gggtcgaggt gagctttcag ccaggcagat
      961 gcacttggcc aggtttctgc ggatgcttct gcgactcgct gatgagtttg gtgtagcagt
     1021 ggtaatcact aatcaggtgg tagctcaagt ggatggagca gcgatgtttg ctgctgatcc
     1081 caaaaaacct attggaggaa atatcatcgc ccatgcatca acaaccagat tgtatctgag
     1141 gaaaggaaga ggggaaacca gaatctgcaa aatctacgac tctccctgtc ttcctgaagc
     1201 tgaagctatg ttcgccatta atgcagatgg agtgggagat gccaaagact gaatcattgg
     1261 gtttttcctc tgttaaaaac cttaagtgct gcagcctaat gagagtgcac tgctccctgg
     1321 ggttctctac aggcctcttc ctgttgtgac tgccaggata aagcttccgg gaaaacagct
     1381 attatatcag cttttctgat ggtataaaca ggagacaggt cagtagtcac aaactgatct
     1441 aaaatgttta ttccttctgt agtgtattaa tctctgtgtg ttttctttgg ttttggagga
     1501 ggggtatgaa gtatctttga catggtgcct taggaatgac ttgggtttaa caagctgtct
     1561 actggacaat cttatgtttc caagagaact aaagctggag agacctgacc cttctctcac
     1621 ttctaaatta atggtaaaat aaaatgcctc agctatgtag caaagggaat gggtctgcac
     1681 agattctttt tttctgtcag taaaactctc aagcaggttt ttaagttgtc tgtctgaatg
     1741 atcttgtgta agggtttggt tatggagtct tgtgccaaac ctactaggcc attagccctt
     1801 caccatctac ctgcttggtc tttcattgct aagactaact caagataatc ctagagtctt
     1861 aaagcatttc aggccagtgt ggtgtcttgc gcctgtactc ccagcacttt gggaggccga
     1921 ggcaggtgga tcgcttgagc caggagtttt aagtccagct tggccaagat ggtgaaatcc
     1981 catctctaca aaaaatgcag aacttaatct ggacacactg ttacacgtgc ctgtagtccc
     2041 agctactcta tagcctgagg tgggagaatc acttaagcct ggaaggtgga agttgcagtg
     2101 agtcgagatt gcactgctgc attccagcca gggtgacaga gtgagaccat gtttcaaaca
     2161 agaaacattt cagagggcaa gtaaacagat ttgattgtga ggcttctaat aaagtagtta
     2221 ttagtagtg
//
```

Revised: October 24, 2001.



☐ 1: NM_079844. Drosophila melano...[gi:17864107] Related Sequences, Protein, PubMed, Taxonomy, LinkOut

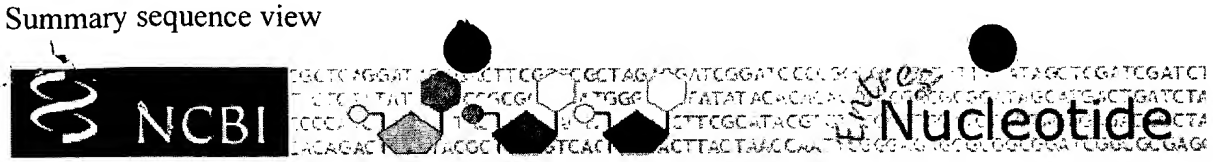
```
LOCUS       NM_079844               1249 bp    mRNA    linear   INV 15-DEC-2001
DEFINITION  Drosophila melanogaster Rad51-like (Rad51), mRNA.
ACCESSION   NM_079844
VERSION     NM_079844.1  GI:17864107
KEYWORDS    .
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (sites)
  AUTHORS   Akaboshi,E., Inoue,Y. and Ryo,H.
  TITLE     Cloning of the cDNA and genomic DNA that correspond to the
            recA-like gene of Drosophila melanogaster
  JOURNAL   Jpn. J. Genet. 69 (6), 663-670 (1994)
  MEDLINE   95161094
REFERENCE   2  (bases 1 to 1249)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor,G.L.,
            Abril,J.F., Agbayani,A., An,H.J., Andrews-Pfannkoch,C., Baldwin,D.,
            Ballew,R.M., Basu,A., Baxendale,J., Bayraktaroglu,L., Beasley,E.M.,
            Beeson,K.Y., Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S.,
            Borkova,D., Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P.,
            Burtis,K.C., Busam,D.A., Butler,H., Cadieu,E., Center,A.,
            Chandra,I., Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B.,
            Davies,P., de Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I.,
            Dietz,S.M., Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S.,
            Dunkov,B.C., Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C.,
            Ferriera,S., Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., WoodageT,
```

Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F., Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H., Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O., Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
TITLE The genome sequence of Drosophila melanogaster
JOURNAL Science 287 (5461), 2185-2195 (2000)
MEDLINE 20196006
PUBMED 10731132
REFERENCE 3 (sites)
AUTHORS Akaboshi,E. and Inoue,Y.
TITLE Cloning of the Drosophila RecA-like cDNA
JOURNAL Unpublished (1994)
REFERENCE 4 (bases 1 to 1249)
AUTHORS Akaboshi,E.
JOURNAL Unpublished (1995)
COMMENT PROVISIONAL REFSEQ: This record has not yet been subject to final NCBI review. The reference sequence was derived from D17726.1.
FEATURES Location/Qualifiers
source 1..1249
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="3"
/map="99D5-99D7"
/cell_line="SICII (GM1)"
gene 1..1249
/gene="Rad51"
/note="CG7948; RAD51; DmRad51; Rad51dm; DMR/DroRAD51; DMR1; RA51; CT6389"
/db_xref="FLYBASE:FBgn0011700;"
/db_xref="LocusID:43577"
CDS 63..1073
/gene="Rad51"
/EC_number="3.6.1.3"
/codon_start=1
/db_xref="FLYBASE:FBgn0011700;"
/db_xref="LocusID:43577"
/product="Rad51-like"
/protein_id="NP_524583.1"
/db_xref="GI:17864108"
/translation="MEKLTNVQAQQEEEEEEGPLSVTKLIGGSITAKDIKLLQQASLH
TVESVANATKKQLMAIPGLGGGKVEQIITEANKLVPLGFLSARTFYQMRADVVQLSTG
SKELDKLLGGGIETGSITEIFGEFRCGKTQLCHTLAVTCQLPISQKGGEGKCMYIDTE
NTFRPERLAAIAQRYKLNESEVLDNVAFTRAHNSDQQTKLIQMAAGMLFESRYALLIV
DSAMALYRSDYIGRGELAARQNHLGLFLRMLQRLADEFGVAVVITNQVTASLDGAPGM
FDAKKPIGGHIMAHSSTTRLYLRKGKGETRICKIYDSPCLPESEAMFAILPDGIGDAR
ES"
misc_feature 432..455
/note="nucleotide binding consensus sequence (A box)"
misc_feature 705..719
/note="nucleotide binding consensus sequence (B box)"
BASE COUNT 322 a 315 c 338 g 274 t
ORIGIN

```
        1 gtacaattta aataacgaaa ctcgaaacaa caaatagcca aatttaactt cgaaactgtc
       61 aaatggagaa gctaacgaat gttcaggcac agcaggaaga ggaggaggaa gaaggtccac
      121 tcagcgtgac taagttaata ggcggcagca tcacggccaa ggacatcaag ctgctccagc
      181 aggccagtct gcacaccgtg gagtcggtgg ccaatgccac caagaagcaa ctgatggcca
      241 ttcccggctt gggcggcggc aaggtggagc agatcatcac ggaggccaac aaactggtgc
      301 ctctgggctt ccttagtgcc cgcaccttct atcaaatgcg tgccgatgtt gtgcagctga
      361 gcacgggctc caaggagctg gacaaactgc tgggcggcgg cattgagacg ggatccatta
      421 ccgagatctt cggcgagttc cgctgcggaa agacgcaatt gtgccacact ctggcagtaa
      481 cctgccagct gcccatcagc cagaagggcg gcgagggcaa gtgcatgtac atcgacacgg
      541 agaacacctt ccgtccggag cgtttggcag ccatcgcgca aaggtacaaa ctgaatgaat
      601 ccgaggtgct ggacaatgtg gccttcaccc gtgcccacaa ctcagatcag cagaccaagc
```

```
     661 tcatccagat ggcggcgggc atgctctttg agtccagata cgctttgctg attgtggaca
     721 gtgccatggc gctctacaga tccgattata ttggtcgcgg ggagctggcc gccaggcaaa
     781 accatttggg cttatttctg cgcatgcttc aacgcctggc cgatgagttc ggagtggctg
     841 tggtaattac taaccaggtt actgcctcgc tggacggcgc acccggcatg tttgatgcca
     901 agaagcccat tggcgggcac atcatggccc actcctcaac cacgcggctg tatctgcgca
     961 agggtaaagg cgaaacccgc atctgcaaga tctacgactc gccctgtttg ccggaatcgg
    1021 aggccatgtt cgccattctg ccggatggaa taggagacgc cagggagagc taattgtgct
    1081 cacttatagg ttaattaatg ctaggaatag agctccctaa ctttctaatg attacttatc
    1141 cctagactaa gaaataaact tcgtagattg ttttttttt ttgttttaat gtttacttag
    1201 aatgtttagg ataacgacga aataaaacag ccgttgcgaa ttgtatgta
//
```

Revised: October 24, 2001.



FASTA view

Sequence feature view of the region: (gi|6598365:48046-48220, 48625-48744, 48827-48860, 48943-49013, 49092-49322, 49436-49560, 49646-49712, 49809-49984)

LOCUS AC002387 999 bp DNA PLN 05-APR-200
DEFINITION CDS from: Arabidopsis thaliana chromosome II section 242 of the complete sequence. Sequence from clones T14P1, F4L23.
ACCESSION AC002387
VERSION AC002387.2
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheo Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis
REFERENCE 1 (bases 48046 to 49984)
AUTHORS Lin,X., Kaul,S., Rounsley,S.D., Shea,T.P., Benito,M.-I., Tow Fujii,C.Y., Mason,T.M., Bowman,C.L., Barnstead,M.E., Feldblyum,T.V., Buell,C.R., Ketchum,K.A., Lee,J.J., Ronning, Koo,H., Moffat,K.S., Cronin,L.A., Shen,M., VanAken,S.E., Uma Tallon,L.J., Gill,J.E., Adams,M.D., Carrera,A.J., Creasy,T.H Goodman,H.M., Somerville,C.R., Copenhaver,G.P., Preuss,D., Nierman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C. Venter,J.C.
TITLE Sequence and analysis of chromosome 2 of the plant Arabidops thaliana
JOURNAL Nature 402 (6763), 761-768 (1999)
MEDLINE 20083487
PUBMED 10617197
REFERENCE 2 (bases 48046 to 49984)
AUTHORS Lin,X.
TITLE Direct Submission
JOURNAL Submitted (09-MAR-2000) The Institute for Genomic Research, Medical Center Dr., Rockville, MD 20850, USA
COMMENT On Dec 17, 1999 this sequence version replaced gi:2583106. The sequence and annotation of chromosome 2 were merged from of the individual clones on this chromosome after removing overlaps. For detailed information, please see the TIGR web (http://www.tigr.org/tdb/at/at.html).

Genes were identified by a combination of three methods: Gen prediction programs including GRAIL (ftp://arthur.epm.ornl.gov/pub/xgrail), Genefinder (Phil Gre University of Washington), Genscan (Chris Burge, http://gnomic.stanford.edu/GENSCANW.html), and NetPlantGene (http://www.cbs.dtu.dk/services/NetGene2/), searches of the complete sequence against a peptide database and plant EST databases at TIGR, and manual curations based on those analy Annotated genes are named to indicate the level of evidence their annotation. Genes with similarity to other proteins ar after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' pr Genes without protein or EST similarity, that are predicted or more gene prediction programs over most of their length a annotated as 'hypothetical' proteins. Genes encoding tRNAs a

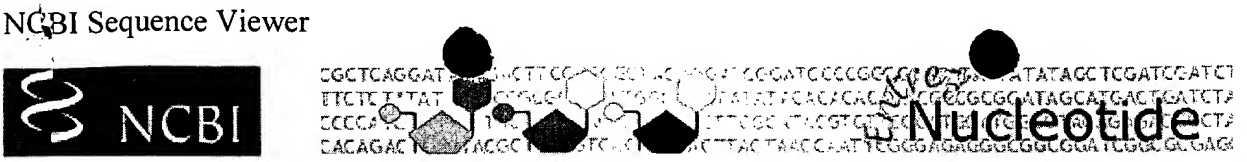
predicted by tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats w identified by repeatmasker (Arian Smit, http://ftp.genome.washington.edu/RM/RepeatMasker.html). Gene numbered from the top to bottom of the chromosome.

We thank the CSHL/WashU/ABI consortium for sequencing BAC cl F6P23, F5J6, T17A5, and T13L16, the ESSA group for sequencin F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Ri and Satoshi Tabata for helpful assistance. In addition, we w like to thank the TIGR Bioinformatics Department, especially Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Peterson, Michael Holmes, and Delwood Richardson for softwar database support.

This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.

Address all correspondence to: at@tigr.org.

```
FEATURES             Location/Qualifiers
     misc_feature    complement(<1..>999)
                     /note="Sequence from clone F4L23"
     mRNA            <1..>999
                     /gene="At2g45280"
     gene            <1..>999
                     /gene="At2g45280"
                     /note="F4L23.21"
     CDS             1..999
                     /gene="At2g45280"
                     /codon_start=1
                     /product="putative RAD51C-like DNA repair protein"
                     /protein_id="AAB82635.1"
                     /db_xref="GI:2583126"
                     /translation="MISFGRRKSPAIEETSLATSVMEAWRLPLSPSIRGKL
                     LSSIASVSSSDLARAKNAWDMLHEEESLPRITTSCSDLDNILGGGISCRDV
                     GIGKTQIGIQLSVNVQIPRECGGLGGKAIYIDTEGSFMVERALQIAEACVE
                     YMHKHFQANQVQMKPEDILENIFYFRVCSYTEQIALVNHLEKFISENKDVV
                     FHFRQDYDDLAQRTRVLSEMALKFMKLAKKFSLAVVLLNQVTTKFSEGSFQ
                     SWSHSCTNRVILYWNGDERYAYIDKSPSLPSASASYTVTSRGLRNSSSSSK
BASE COUNT      272 a    203 c    243 g    281 t
ORIGIN
        1 atgatttcat ttgggcggcg taaatcgccg gcgattgaag aaacttcact cgcgact
       61 gtcatggagg catggaggtt accgttatcg ccttcgatta gaggaaaact gatatcg
      121 ggttatactt gtctgtcttc gattgcttcc gtctcttctt ctgatctcgc tcgagca
      181 aacgcttggg atatgcttca cgaggaggag tctttgccgc gtattactac atcttgc
      241 gatcttgata acattttggg cggtggaatt agctgtaggg atgttacaga gattggt
      301 gtaccaggga ttggcaagac tcagattggg atccagctct ctgtgaatgt tcagatt
      361 cgtgagtgtg gtggtcttgg agggaaagct atatatatcg atacagaagg tagcttc
      421 gtggagcgtg ctttacagat agcagaagct tgtgtagagg acatggaaga atacaca
      481 tacatgcata aacattttca agcaaatcaa gtacaaatga aaccagaaga tatctta
      541 aacatattct acttccgtgt ctgcagttac accgagcaaa tcgcattggt caatcat
      601 gaaaagttca tctctgaaaa caaagatgta gttgtaatcg tagacagtat caccttt
      661 ttccgtcagg actatgatga cttagcccag aggacacgag tgctcagcga aatggct
      721 aagttcatga agcttgccaa aaagttctca cttgcggtcg tgttactaaa ccaggtg
      781 acaaagttta gtgaaggctc gtttcaacta gcgcttgctt taggcgatag ctggtct
      841 tcgtgcacca accgagtcat tctgtattgg aatggtgatg agcgttacgc atatatc
      901 aagtcccctt cacttccttc agcttcggct tcatacactg taaccagtag aggtcta
      961 aactcatcct cgagtagcaa gcgagtcaag atgatgtaa
//
```

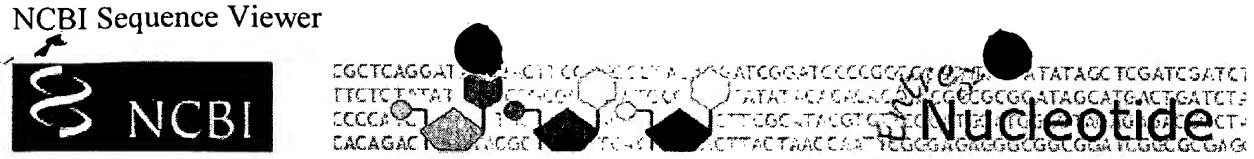


☐ 1: AF029669. Homo sapiens Rad5...[gi:2909800] Related Sequences, OMIM, Protein, PubMed, Taxonomy, LinkOut

```
LOCUS       AF029669                1295 bp    mRNA    linear   PRI 24-FEB-1998
DEFINITION  Homo sapiens Rad51C (RAD51C) mRNA, complete cds.
ACCESSION   AF029669
VERSION     AF029669.1  GI:2909800
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1295)
  AUTHORS   Dosanjh,M.K., Collins,D.W., Fan,W., Lennon,G.G., Albala,J.S.,
            Shen,Z. and Schild,D.
  TITLE     Isolation and characterization of RAD51C, a new human member of the
            RAD51 family of related genes
  JOURNAL   Nucleic Acids Res. 26 (5), 1179-1184 (1998)
  MEDLINE   98136197
REFERENCE   2  (bases 1 to 1295)
  AUTHORS   Schild,D., Collins,D.W. and Dosanjh,M.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-1997) Life Sciences, Lawrence Berkeley National
            Laboratory, Ms. 70A-1118, Berkeley, CA 94720, USA
FEATURES             Location/Qualifiers
     source          1..1295
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17q; 413.6cR from the top"
                     /cell_type="leukocyte (mixed population)"
     gene            1..1295
                     /gene="RAD51C"
     CDS             43..1173
                     /gene="RAD51C"
                     /note="member of the Rad51 family of related proteins"
                     /codon_start=1
                     /product="Rad51C"
                     /protein_id="AAC39604.1"
                     /db_xref="GI:2909801"
                     /translation="MRGKTFRFEMQRDLVSFPLSPAVRVKLVSAGFQTAEELLEVKPS
                     ELSKEVGISKAEALETLQIIRRECLTNKPRYAGTSESHKKCTALELLEQEHTQGFIIT
                     FCSALDDILGGGVPLMKTTEICGAPGVGKTQLCMQLAVDVQIPECFGGVAGEAVFIDT
                     EGSFMVDRVVDLATACIQHLQLIAEKHKGEEHRKALEDFTLDNILSHIYYFRCRDYTE
                     LLAQVYLLPDFLSEHSKVRLVIVDGIAFPFRHDLDDLSLRTRLLNGLAQQMISLANNH
                     RLAVILTNQMTTKIDRNQALLVPALGESWGHAATIRLIFHWDRKQRLATLYKSPSQKE
                     CTVLFQIKPQGFRDTVVTSACSLQTEGSLSTRKRSRDPEEEL"
BASE COUNT      392 a    249 c    308 g    346 t
ORIGIN
        1 gtgcggagtt tggctgctcc ggggttagca ggtgagcctg cgatgcgcgg gaagacgttc
       61 cgctttgaaa tgcagcggga tttggtgagt ttcccgctgt ctccagcggt gcgggtgaag
      121 ctggtgtctg cggggttcca gactgctgag gaactcctag aggtgaaacc ctccgagctt
      181 agcaaagaag ttgggatatc taaagcagaa gccttagaaa ctctgcaaat tatcagaaga
```

```
      241 gaatgtctca caaataaacc aagatatgct ggtacatctg agtcacacaa gaagtgtaca
      301 gcactggaac ttcttgagca ggagcatacc cagggcttca taatcacctt ctgttcagca
      361 ctagatgata ttcttggggg tggagtgccc ttaatgaaaa caacagaaat ttgtggtgca
      421 ccaggtgttg gaaaaacaca attatgtatg cagttggcag tagatgtgca gataccagaa
      481 tgttttggag gagtggcagg tgaagcagtt tttattgata cagagggaag ttttatggtt
      541 gatagagtgg tagaccttgc tactgcctgc attcagcacc ttcagcttat agcagaaaaa
      601 cacaagggag aggaacaccg aaaagctttg gaggatttca ctcttgataa tattctttct
      661 catatttatt attttcgctg tcgtgactac acagagttac tggcacaagt ttatcttctt
      721 ccagatttcc tttcagaaca ctcaaaggtt cgactagtga tagtggatgg tattgctttt
      781 ccatttcgtc atgacctaga tgacctgtct cttcgtactc ggttattaaa tggcctagcc
      841 cagcaaatga tcagccttgc aaataatcac agattagctg taattttaac caatcagatg
      901 acaacaaaga ttgatagaaa tcaggccttg cttgttcctg cattagggga aagttgggga
      961 catgctgcta caatacggct aatctttcat tgggaccgaa agcaaaggtt ggcaacattg
     1021 tacaagtcac ccagccagaa ggaatgcaca gtactgtttc aaatcaaacc tcagggattt
     1081 agagatactg ttgttacttc tgcatgttca ttgcaaacag aaggttcctt gagcacccgg
     1141 aaacggtcac gagacccaga ggaagaatta taacccagaa acaaatctca aagtgtacaa
     1201 atttattgat gttgtgaaat caatgtgtac aagtggactt gttaccttaa agtataaata
     1261 aacacactat ggcatgaatg aaaaaaaaaa aaaaa
//
```

Revised: October 24, 2001.



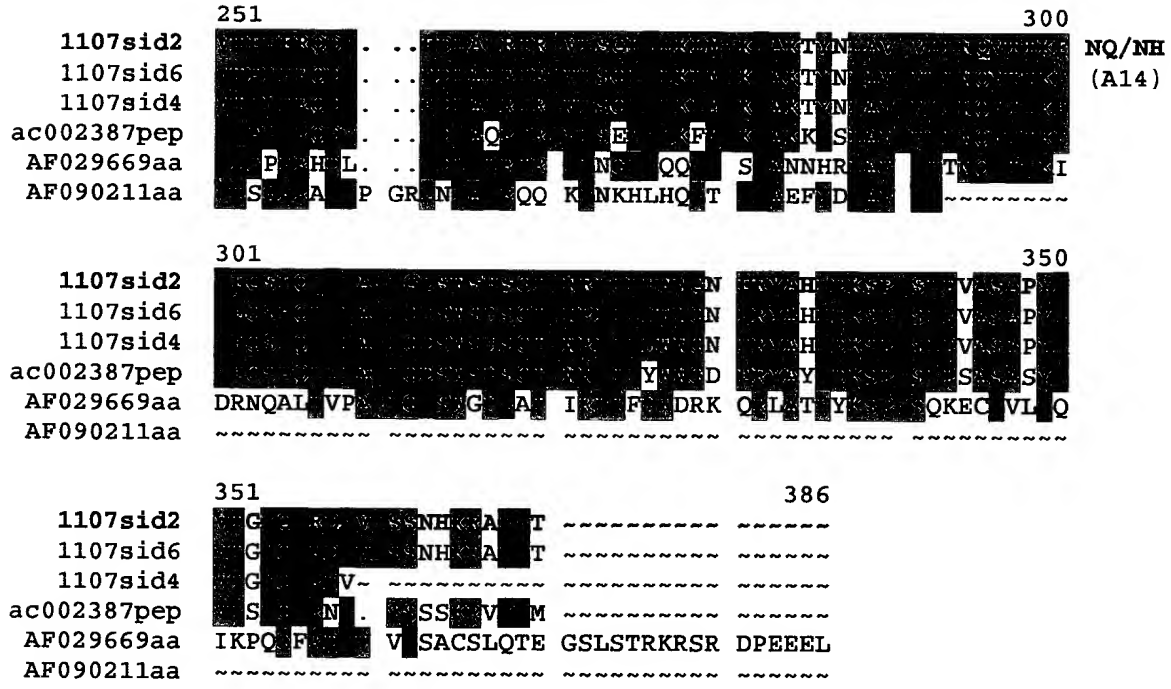
1: U84138. Homo sapiens DNA ...[gi:2801404] Related Sequences, OMIM, Protein, PubMed, Taxonomy, LinkOut

```
LOCUS       HSU84138                 1764 bp    mRNA    linear   PRI 05-MAY-1998
DEFINITION  Homo sapiens DNA repair protein RAD51B mRNA, complete cds.
ACCESSION   U84138
VERSION     U84138.1  GI:2801404
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1764)
  AUTHORS   Albala,J.S., Thelen,M.P., Prange,C., Fan,W., Christensen,M.,
            Thompson,L.H. and Lennon,G.G.
  TITLE     Identification of a novel human RAD51 homolog, RAD51B
  JOURNAL   Genomics 46 (3), 476-479 (1997)
  MEDLINE   98110585
REFERENCE   2  (bases 1 to 1764)
  AUTHORS   Albala,J.S., Prange,C.K., Fan,W., Christensen,M., Thelen,M. and
            Lennon,G.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JAN-1997) Biology and Biotechnology Research Program,
            Lawrence Livermore National Laboratory, 7000 East Avenue,
            Livermore, CA 94550, USA
FEATURES             Location/Qualifiers
     source          1..1764
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /map="14q23-q24.2"
     5'UTR           1..54
     CDS             55..1107
                     /codon_start=1
                     /product="DNA repair protein RAD51B"
                     /protein_id="AAC39723.1"
                     /db_xref="GI:2801405"
                     /translation="MGSKKLKRVGLSQELCDRLSRHQILTCQDFLCLSPLELMKVTGL
                     SYRGVHELLCMVSRACAPKMQTAYGIKAQRSADFSPAFLSTTLSALDEALHGGVACGS
                     LTEITGPPGCGKTQFCIMMSILATLPTNMGGLEGAVVYIDTESAFSAERLVEIAESRF
                     PRYFNTEEKLLLTSSKVHLYRELTCDEVLQRIESLEEEIISKGIKLVILDSVASVVRK
                     EFDAQLQGNLKERNKFLAREASSLKYLAEEFSIPVILTNQITTHLSGALASQADLVSP
                     ADDLSLSEGTSGSSCVIAALGNTWSHSVNTRLILQYLDSERRQILIAKSPLAPFTSFV
                     YTIKEEGLVLQAYGNS"
     3'UTR           1108..1764
     polyA_signal    1727..1732
BASE COUNT      513 a    361 c    387 g    503 t
ORIGIN
        1 gggaaactgt gtaaagggtg gggaaacttg aaagttggat gctgcagacc cggcatgggt
       61 agcaagaaac taaaacgagt gggtttatca caagagctgt gtgaccgtct gagtagacat
      121 cagatcctta cctgtcagga ctttttatgt ctttccccac tggagcttat gaaggtgact
      181 ggtctgagtt atcgaggtgt ccatgaactt ctatgtatgg tcagcagggc ctgtgcccca
      241 aagatgcaaa cggcttatgg gataaaagca caaaggtctg ctgatttctc accagcattc
```

```
     301 ttatctacta ccctttctgc tttggacgaa gccctgcatg gtggtgtggc ttgtggatcc
     361 ctcacagaga ttacaggtcc accaggttgt ggaaaaactc agttttgtat aatgatgagc
     421 attttggcta cattacccac caacatggga ggattagaag gagctgtggt gtacattgac
     481 acagagtctg catttagtgc tgaaagactg gttgaaatag cagaatcccg ttttcccaga
     541 tattttaaca ctgaagaaaa gttacttttg acaagtagta aagttcatct ttatcgggaa
     601 ctcacctgtg atgaagttct acaaaggatt gaatctttgg aagaagaaat tatctcaaaa
     661 ggaattaaac ttgtgattct tgactctgtt gcttctgtgg tcagaaagga gtttgatgca
     721 caacttcaag gcaatctcaa agaaagaaac aagttcttgg caagagaggc atcctccttg
     781 aagtatttgg ctgaggagtt ttcaatccca gttatcttga cgaatcagat tacaacccat
     841 ctgagtggag ccctggcttc tcaggcagac ctggtgtctc cagctgatga tttgtccctg
     901 tctgaaggca cttctggatc cagctgtgtg atagccgcac taggaaatac ctggagtcac
     961 agtgtgaata cccggctgat cctccagtac cttgattcag agagaagaca gattcttatt
    1021 gccaagtccc ctctggctcc cttcacctca tttgtctaca ccatcaagga ggaaggcctg
    1081 gttcttcaag cctatggaaa ttcctagaga cagataaatg tgcaaacctg ttcatcttgc
    1141 caagaaaaat ccgcttttt gccacagaaa caaaatattg ggaaagagtc ttgtggtgaa
    1201 acacccatcg ttctttgcta aaacatttgg ttgctactgt gtagactcag cttaagtcat
    1261 ggaattctag aggatgtatc tcacaagtag gatcaagaac aagcccaaca gtaatctgca
    1321 tcataagctg atttgatacc atggcactga caatgggcac tgatttgata ccatggcact
    1381 gacatgggca cacagggaac aggaaatggg aatgagagca agggttgggt tgtgttcgtg
    1441 gaacacatag gttttttttt tttaactttc tctttctaaa atatttcatt ttgatggagg
    1501 tgaaatttat ataagatgaa attaaccatt ttaaagtaaa caattccgtg gcaactagat
    1561 atcatgatgt gcaaccagca tctctgtcta gttccaaata ttttcatcac cccaaaagca
    1621 agacccataa ccattatgca agtgttccta tttccccctc ctcccagctc ctggaaaccc
    1681 accaatctac tttgttgcta tggctttacc tattctggat atttcatata aatggaatca
    1741 tatagtgtca taaaaaaaaa aaaa
//
```

Revised: October 24, 2001.

```
!!AA_MULTIPLE_ALIGNMENT 1.0
PileUp of: @/tmp/64080830.list

 Symbol comparison table: genrundata:blosum62.cmp  CompCheck: 1102

                  GapWeight: 8      GapLengthWeight: 2


 1107sid2         corn Rad51C SEQ ID NO: 2 encoded by SEQ ID NO: 1 (elected)
 1107sid6         corn Rad51C SEQ ID NO: 6 encoded by SEQ ID NO: 5
 1107sid4         corn Rad51C SEQ ID NO: 4 encoded by SEQ ID NO: 3
 ac002387pep      A. thaliani Rad51C protein encoded by GB AC002387 (F4L23.21)
 AF029669aa       Human Rad51C protein encoded by GB AF029669
 AF090211aa       Archaea RadA/Rad51C protein encoded by partial cds GB AF090211

Format: [redacted];
Conserved motifs underlined, references to specification or IDS submission

            1                                                   50
   1107sid2 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
   1107sid6 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
   1107sid4 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
ac002387pep ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~MISFGRRKS PAIEETSLAT
 AF029669aa MRGKTFRFEM QRDLVSFPLS PAVRVKLVSA GFQTAEELLE VKPSELSKEV
 AF090211aa ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

            51                                                  100
   1107sid2 ~~~~~~~~~~ ~~~~~~~~~~ MGDQS.GSRN GPQQKY[redacted]
   1107sid6 ~~~~~~~~~~ ~~~~~~~~~~ MSDQS.GSRN GPQQKY[redacted]
   1107sid4 ~~~~~~~~~~ ~~~~~~~~~~ MSDQS.GSRN GPQQKY[redacted]
ac002387pep SVMEAWRLPL SPSIRGKLIS A[redacted]
 AF029669aa GISKAEALET LQIIRRECLT NKPRYA[redacted]
 AF090211aa ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

            101                                                 150
   1107sid2 L[redacted]                                          ATP-binding
   1107sid6 L[redacted]                                          (Example 4)
   1107sid4 L[redacted]                                          (& A13)
ac002387pep E[redacted]
 AF029669aa H[redacted]
 AF090211aa ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

            151                                                 200
   1107sid2 [redacted]
   1107sid6 [redacted]
   1107sid4 [redacted]
ac002387pep [redacted]
 AF029669aa [redacted]
 AF090211aa [redacted]

            201                                                 250
   1107sid2 [redacted]                                           ATP-binding
   1107sid6 [redacted]                                           (A13, A14)
   1107sid4 [redacted]
ac002387pep [redacted]
 AF029669aa [redacted]
 AF090211aa [redacted]
```


251
300
1107sid2
1107sid6
1107sid4
ac002387pep
AF029669aa
AF090211aa
NQ/NH
(A14)
301
350
1107sid2
1107sid6
1107sid4
ac002387pep
AF029669aa
AF090211aa
351
386
1107sid2
1107sid6
1107sid4
ac002387pep
AF029669aa
AF090211aa

Moderate stripping solution

200 mM Tris·Cl, pH 7.0
0.1× SSC (*APPENDIX 2*)
0.1% (w/v) SDS

Nucleotide mix

2.5 mM ATP
2.5 mM CTP
2.5 mM GTP
20 mM Tris·Cl, pH 7.5
Store at –20°C

COMMENTARY

Background Information

Hybridization between complementary polynucleotides was implicit in the Watson-Crick model for DNA structure and was initially exploited, via renaturation kinetics, as a means of studying genome complexity. In these early applications, the two hybridizing molecules were both in solution—an approach that is still utilized in "modern" techniques such nuclease protection transcript mapping (*UNITS 4.6 & 4.7*) and oligonucleotide-directed mutagenesis (Chapter 8). The innovative idea of immobilizing one hybridizing molecule on a solid support was first proposed by Denhardt (1966) and led to methods for identification of specific sequences in genomic DNA (dot blotting; *UNIT 2.9B*) and recombinant clones (*UNITS 6.3 & 6.4*). A second dimension was subsequently introduced by Southern (1975), who showed how DNA molecules contained in an electrophoresis gel could be transferred to a membrane (*UNIT 2.9A*), enabling genetic information relating to individual restriction fragments to be obtained by hybridization analysis.

Since the pioneering work of Denhardt and Southern, advances in membrane hybridization have been technical rather than conceptual. As reviewed by Dyson (1991), the detailed protocols have become more sophisticated, largely because of advances in understanding of the factors that influence hybrid stability and hybridization rate.

Hybrid stability is expressed as the melting temperature or T_m, which is the temperature at which the probe dissociates from the target DNA. For DNA-DNA hybrids, the T_m can be approximated from the equation of Meinkoth and Wahl (1984):

$$T_m = 81.5°C + 16.6 (\log M) + 0.41(\%GC) - 0.61 (\%form) - 500/L$$

and for RNA-DNA hybrids from the equation of Casey and Davidson (1977):

$$T_m = 79.8°C + 18.5(\log M) + 0.58(\%GC) - 11.8(\%GC)^2 - 0.56(\%form) - 820/L$$

where M is the molarity of monovalent cations, %GC is the percentage of guanosine and cytosine nucleotides in the DNA, %form is the percentage of formamide in the hybridization solution, and L is the length of the hybrid in base pairs. The practical considerations that arise from these two equations are summarized Table 2.10.2A.

The second important consideration in hybridization analysis is the rate at which the hybrid is formed. Hybrid formation cannot occur until complementary regions of the two molecules become aligned, which occurs purely by chance; however, once a short nucleating region of the duplex has formed, the remaining sequences base-pair relatively rapidly. The rate at which the probe "finds" the target, which is influenced by a number of factors (Table 2.10.2B), is therefore the limiting step in hybrid formation (Britten and Davidson, 1985). However, in practical terms hybridization rate is less important than hybrid stability, as in most protocols hybridization is allowed to proceed for so long that factors influencing rate become immaterial.

Critical Parameters

To be successful, a hybridization experiment must meet two criteria:

(1) *Sensitivity.* Sufficient probe DNA must anneal to the target to produce a detectable signal after autoradiography.

(2) *Specificity.* After the last wash, the probe must be attached only to the desired target sequence (or, with heterologous probing, family of sequences).

Parameters influencing these two criteria are considered in turn, followed by other miscellaneous factors that affect hybridization.

Factors influencing sensitivity

The sensitivity of hybridization analysis is determined by how many labeled probe molecules attach to the target DNA. The greater the number of labeled probe molecules that anneal, the greater the intensity of the hybridization signal seen after autoradiography.

Probe specific activity. Of the various factors that influence sensitivity, the one that most frequently causes problems is the specific ac-

tivity of the probe. Modern labeling procedures, whether nick translation, random oligonucleotide priming (*UNIT 3.5*), or in vitro RNA synthesis (alternate protocol), routinely provide probes with a specific activity of $>10^8$ dpm/μg. This is the minimum specific activity that should be used in hybridization analysis of genomic DNA, even if the target sequences are multicopy. If the specific activity is $<10^8$ dpm/μg, hybridization signals will be weak or possibly undetectable, and no amount of adjusting the hybridization conditions will compensate. If there is a problem in obtaining a specific activity of $>10^8$ dpm/μg, it is important to troubleshoot the labeling protocol before attempting to use the probe in hybridization analysis.

If the probe is labeled to 10^8 to 10^9 dpm/μg, it will be able to detect as little as 0.5 pg of target DNA. Exactly what this means depends on the size of the genome being studied and the copy number of the target sequence. For human genomic DNA, 0.5 pg of a single-copy sequence 500 bp in length corresponds to 3.3 μg of total DNA. This is therefore the minimum amount of human DNA that should be used in a dot blot or Southern transfer if a single-copy gene is being sought.

Amount of target DNA. There is, however, a second argument that dictates that rather more than 3.3 μg of DNA should be loaded with each dot or Southern blot. During hybridization, genuine target sequences (100% homologous to the probe) and heterologous target sequences (related but not identical to the probe) compete with one another, with the homologous reactions always predominant. Ideally this competition should be maintained until the end of the incubation period so that maximum discrimination is seen between homologous and heterologous signals. This occurs only if the membrane-bound DNA is in excess, so that target sequences are continually competing for the available probe (Anderson and Young, 1985). If the probe is in excess then the homologous reaction may reach completion (i.e., all genuine target sites become filled) before the end of the incubation, leaving a period when only

Table 2.10.2 Factors Influencing Hybrid Stability and Hybridization Rate[a]

Factor	Influence
A. Hybrid stability[b]	
Ionic strength	T_m increases 16.6°C for each 10-fold increase in monovalent cations between 0.01 and 0.40 M NaCl
Base composition	AT base pairs are less stable than GC base pairs in aqueous solutions containing NaCl
Destabilizing agents	Each 1% of formamide reduces the T_m by about 0.6°C for a DNA-DNA hybrid. 6 M urea reduces the T_m by about 30°C
Mismatched base pairs	T_m is reduced by 1°C for each 1% of mismatching
Duplex length	Negligible effect with probes >500 bp
B. Hybridization rate[b]	
Temperature	Maximum rate occurs at 20-25°C below T_m for DNA-DNA hybrids, 10-15°C below T_m for DNA-RNA hybrids
Ionic strength	Optimal hybridization rate at 1.5 M Na^+
Destabilizing agents	50% formamide has no effect, but higher or lower concentrations reduce the hybridization rate
Mismatched base pairs	Each 10% of mismatching reduces the hybridization rate by a factor of two
Duplex length	Hybridization rate is directly proportional to duplex length
Viscosity	Increased viscosity increases the rate of membrane hybridization; 10% dextran sulfate increases rate by factor of ten
Probe complexity	Repetitive sequences increase the hybridization rate
Base composition	Little effect
pH	Little effect between pH 5.0 and pH 9.0

[a]This table is based on Brown (1991) with permission from BIOS Scientific Publishers.

[b]There have been relatively few studies of the factors influencing membrane hybridization. In several instances extrapolations are made from what is known about solution hybridization. This is probably reliable for hybrid stability, less so for hybridization rate.

heterologous hybridization occurring and during which discrimination between the homologous and heterologous signals becomes reduced. The problem is more significant with a double-stranded rather than a single-stranded probe, as with double-stranded probe reannealing between the two probe polynucleotides gradually reduces the effective probe concentration to such an extent that it always becomes limiting towards the end of the incubation.

In practical terms it is difficult to ensure that the membrane-bound DNA is in excess. The important factor is not just the absolute amount of DNA (which is dependent on the efficiency of immobilization and how many times the membrane has been reprobed) but also the proportion of the DNA that is composed of sequences (homologous and heterologous) able to hybridize to the probe. Rather than attempting complex calculations whose results may have factor-of-ten errors, it is advisable simply to blot as much DNA as possible: 10 µg is sufficient with most genomes. Assuming that the probe is labeled adequately and used at the correct concentration in the hybridization solution, a clear result will be obtained after autoradiography for a few hours with a simple genome (e.g., yeast DNA) or a few days with a more complex one (e.g., human DNA).

Labels other than ^{32}P. The discussion so far has assumed that the probe is labeled with ^{32}P. The lower emission energy of ^{35}S results in reduced sensitivity, and this isotope is in general unsuitable for hybridization analysis of genomic DNA. ^{35}S can be used only if the blotted DNA is exceptionally noncomplex (e.g., restricted plasmid DNA), or if the DNA is highly concentrated (e.g., colony and plaque blots; *UNIT 6.3*). Note that a membrane hybridized to a ^{35}S-labeled probe has to be dried before autoradiography, so probe stripping is not possible.

Nonradioactive probes are a more realistic option for hybridization analysis of genomic DNA and are becoming increasingly popular as the problems involved in their use are gradually ironed out. Their advantages include greater safety, the fact that large amounts of probe can be prepared in one batch and stored for years, and the rapidity of the detection protocols. Their main disadvantage is that the sensitivity of most nonradioactive detection systems is lower than that of ^{32}P autoradiography, which means that the blot and hybridization have to be carried out at maximum efficiency if a satisfactory signal is to be seen. For details on hybridization analysis with nonradioactive probes, see *UNITS 3.18 & 3.19* and Mundy et al. (1991).

Using an inert polymer to increase sensitivity. In addition to adjusting the parameters discussed above, an improvement in sensitivity can also be achieved by adding an inert polymer such as 10% (w/v) dextran sulfate (molecular weight 500,000) or 8% (w/v) PEG 6000 to the hybridization solution. Both induce an increase in hybridization signal, 10-fold with a single-stranded probe and as much as 100-fold if the probe is double-stranded (Wahl et al., 1979; Amasino, 1986). The improvement is thought to arise from formation of interlocked meshes of probe molecules, which anneal en masse at target sites. Increased hybridization signals are a major bonus in detecting single-copy sequences in complex genomes, but this must be balanced with the fact that the polymers make the hybridization solutions very viscous and difficult to handle.

Factors influencing specificity

Ensuring specificity in homologous hybridization experiments. The hybridization incubation is carried out in a high-salt solution that promotes base-pairing between probe and target sequences. In 5× SSC, the T_m for genomic DNA with a GC content of 50% is about 96°C. Hybridization is normally carried out at 68°C, so the specificity of the experiment is not determined at this stage. Specificity is the function of post-hybridization washes, the critical parameters being the ionic strength of the final wash solution and the temperature at which this wash is carried out.

The highly stringent wash conditions described in the basic and alternate protocols should destabilize all mismatched heteroduplexes, so that hybridization signals are obtained only from sequences that are perfectly homologous to the probe. For DNA and RNA probes (as opposed to oligonucleotides), problems with lack of specificity after the highly stringent wash occur only if the hybridizing sequences are very GC-rich, resulting in a relatively high T_m. If the high-stringency wash does not remove all nonspecific hybridization, temperature can be increased by a few degrees. The equations above for calculating T_m can be used as a guide for selecting the correct temperature for the final wash, but trial and error is more reliable. Note that a membrane that has been autoradiographed can be rewashed at a higher stringency and put back to expose again, the only limitation being decay of the label and the need for a longer exposure the second time.

Table 2.10.3 High-Salt Solutions Used in Hybridization Analysis

Stock solution	Composition
20× SSC	3.0 M NaCl/0.3 M trisodium citrate
20× SSPE[a]	3.6 M NaCl/0.2 M NaH_2PO_4/0.02 M EDTA, pH 7.7
Phosphate solution[b]	1 M $NaHPO_4$, pH 7.2[c]

[a]SSC may be replaced with the same concentration of SSPE in all protocols.

[b]Prehybridize and hybridize with 0.5 M $NaHPO_4$ (pH 7.2)/1 mM EDTA/7% SDS [or 50% formamide/0.25 M $NaHPO_4$ (pH 7.2)/0.25 M NaCl/1 mM EDTA/7% SDS]; perform moderate-stringency wash in 40 mM $NaHPO_4$ (pH 7.2)/1 mM EDTA/5% SDS; perform high-stringency wash in 40 mM $NaHPO_4$ (pH 7.2)/1 mM EDTA/1% SDS.

[c]Dissolve 134 g $Na_2HPO_4 \cdot 7H_2O$ in 1 liter water, then add 4 ml 85% H_3PO_4. The resulting solution is 1 M Na^+, pH 7.2.

Designing washes for heterologous hybridization. Calculations of T_m become more critical if heterologous probing is being attempted. If the aim is to identify sequences that are merely related, not identical, to the probe (e.g., members of a multigene family, or a similar gene in a second organism), then it is useful to have an idea of the degree of mismatching that will be tolerated by a "moderate-" or "low-" stringency wash. The best way to approach this is to first establish the lowest temperature at which only homologous hybridization occurs with a particular SSC concentration. Then assume that 1% mismatching results in a 1°C decrease in the T_m (Bonner et al., 1973) and reduce the temperature of the final wash accordingly (for example, if sequences with ≥90% similarity with the probe are being sought, decrease the final wash temperature by 10°C). If the desired degree of mismatching results in a wash temperature of <45°C, then it is best to increase the SSC concentration so that a higher temperature can be used. Doubling the SSC concentration results in a ~17°C increase in T_m, so washes at 45°C in 0.1× SSC and 62°C in 0.2× SSC are roughly equivalent. Note that in these extreme cases it may also be necessary to reduce the hybridization temperature to as low as 45°C (aqueous solution) or 32°C (formamide solution).

This approach sometimes works extremely well (as shown when the heterologous targets are eventually sequenced), but the assumption that a 1% degree of mismatching reduces the T_m of a heteroduplex by 1°C is very approximate. Base composition and mismatch distribution influence the actual change in T_m, which can be anything between 0.5° and 1.5°C per 1% mismatch (Hyman et al., 1973). Unfortunately trial and error is the only alternative to the "rational" approach described here.

Other parameters relevant to hybridization analysis

Length of prehybridization and hybridization incubations. The protocols recommend prehybridization for 3 hr with nitrocellulose and 15 min for nylon membranes. Inadequate prehybridization can lead to high backgrounds, so these times should not be reduced. They can, however, be extended without problem.

Hybridizations are usually carried out overnight. This is a rather sloppy aspect of the procedure, because time can have an important influence on the result, especially if, as described above, an excess amount of a single-stranded probe is being used. The difficulties in assigning values to the parameters needed to calculate optimum hybridization time has led to the standard "overnight" incubation, which in fact is suitable for most purposes. The exception is when hybridization is being taken to its limits, for instance in detection of single-copy sequences in human DNA, when longer hybridization times (up to 24 hr) may improve sensitivity if a single-stranded probe is being used. Note that this does not apply to double-stranded probes, as gradual reannealing results in only minimal amounts of a double-stranded probe being free to hybridize after ~8 hr of incubation.

Formamide hybridization buffers. Formamide destabilizes nucleic acid duplexes, reducing the T_m by an average of 0.6°C per 1% formamide for a DNA-DNA hybrid (Meinkoth and Wahl, 1984) and rather less for a DNA-RNA hybrid (Casey and Davidson, 1977; Kafatos et al., 1979). It can be used at 50% concentration in the hybridization solution, reducing the T_m so that the incubation can be carried out at a lower temperature than needed with an aqueous solution. Originally formamide was used with nitrocellulose membranes as a means of prolonging their lifetime, as the

lower hybridization temperatures results in less removal of target DNA from the matrix. More recently formamide has found a second use in reduction of heterologous background hybridization with RNA probes. RNA-DNA hybrids are relatively strong, and heterologous duplexes remain stable even at high temperatures. The destabilizing effect of formamide is therefore utilized to maximize the discrimination between homologous and heterologous hybridization with RNA probes.

Formamide probably confers no major advantage on DNA-DNA hybridization with a nylon membrane. In fact it introduces two problems, the hazardous nature of the chemical itself, and an apparent reduction in hybridization rate. The latter point is controversial (Hutton, 1977), but for equivalent sensitivity a formamide hybridization reaction usually has to incubate for longer than an aqueous one.

Alternatives to SSC. Although SSC has been used in hybridization solutions for many years, there is nothing sacrosanct about the formulation, and other salt solutions can be employed (Table 2.10.3). There is little to choose between these alternatives. SSPE and phosphate solutions have a greater buffering power and may confer an advantage in formamide hybridization solutions. Alternatively, the buffering power of SSC can be increased by adding 0.3% (w/v) tetrasodium pyrophosphate.

Probe length. Probe length has a major influence on the rate of duplex formation in solution hybridization (Wetmur and Davidson, 1986), but the effect is less marked when the target DNA is immobilized. In membrane hybridization a more important factor is the specificity of the probe. The probe should never be too long (>1000 bp), as this increases the chance of heterologous duplexes remaining stable during a high-stringency wash. Neither should the probe contain extensive vector sequences, as these can hybridize to their own target sites, wrecking the specificity of the experiment.

Mechanics of hybridization. Traditionally hybridization has been carried out in plastic bags. This technique is messy, radiochemical spills being almost unavoidable, and can lead to detrimental contact effects if too many membranes are hybridized in a single bag. Hybridization incubators are now available from a number of companies and are recommended as a distinct advance over the plastic bag technology. Rotation of the hybridization tube results in excellent mixing, reducing hot spots caused by bubbles and dust and leading to very evenly hybridized membranes. Good-quality results are possible even when ten or more minigel Southerns are hybridized in a single 8.5 × 3.0–inch tube.

If bags are used, they should be of stiff plastic to prevent the sides collapsing on to the membrane, which will lead to high background. The volume of hybridization solution should be sufficient to fill the bag, and no more than two membranes should be hybridized in each bag.

Troubleshooting

Problems in blotting and hybridization reveal themselves when the autoradiograph is developed. A guide to the commonest problems and how to solve them is given in Table 2.10.4 (based on Dyson, 1991).

A particularly troublesome problem is high background signal across the entire membrane. This is due to the probe attaching to nucleic acid binding sites on the membrane surface, the same sites that bind DNA during the blotting procedure. Prehybridization/hybridization solutions contain reagents that block these sites and hence reduce background hybridization. The most popular blocking agent is Denhardt solution, which contains three polymeric compounds (Ficoll, polyvinylpyrrolidone, and BSA) that compete with nucleic acids for the membrane-binding sites. The formulations used in the basic and alternate protocols also include denatured salmon sperm DNA (any complex DNA that is nonhomologous with the target is acceptable) which also competes with the probe for the membrane sites. Blocking agents are included in the prehybridization solution to give them a head start over the probe. With a nylon membrane, the blocking agents may have to be left out of the hybridization solution, as they can interfere with the probe-target interaction. When the membranes are washed, the Denhardt solution and salmon sperm DNA are replaced with SDS, which acts as a blocking agent at concentrations ≥1%.

Other blocking agents can also be used (Table 2.10.5). With DNA blots, the main alternatives to Denhardt are heparin (Singh and Jones, 1984) and milk powder (BLOTTO; Johnson et al., 1984), although Denhardt is generally more effective, at least with nylon membranes. Note that BLOTTO contains RNases and so can be used only in DNA-DNA hybridizations. With an RNA probe, denatured salmon sperm DNA is sometimes replaced by 100 μg/ml yeast tRNA, which has the advantage that it does not need to be sheared before

Table 2.10.4 Troubleshooting Guide for DNA Blotting and Hybridization Analysis[a]

Problem	Possible cause[b]	Solution
Poor signal	Probe specific activity too low	Check labeling protocol if specific activity is $<10^8$ dpm/μg.
	Inadequate depurination	Check depurination if transfer of DNA >5 kb is poor.
	Inadequate transfer buffer	1. Check that 20× SSC has been used as the transfer solution if small DNA fragments are retained inefficiently when transferring to nitrocellulose. 2. With some brands of nylon membrane, add 2 mM Sarkosyl to the transfer buffer. 3. Try alkaline blotting to a positively charged nylon membrane.
	Not enough target DNA	Refer to text for recommendations regarding amount of target DNA to load per blot.
	Poor immobilization of DNA	See recommendations in *UNIT 2.9A* commentary.
	Transfer time too short	See recommendations in *UNIT 2.9A* commentary.
	Inefficient transfer system	Consider vacuum blotting as an alternative to capillary transfer.
	Probe concentration too low	1. Check that the correct amount of DNA has been used in the labeling reaction. 2. Check recovery of the probe after removal of unincorporated nucleotides. 3. Use 10% dextran sulfate in the hybridization solution. 4. Change to a single-stranded probe, as reannealing of a double-stranded probe reduces its effective concentration to zero after hybridization for 8 hr.
	Incomplete denaturation of probe	Denature as described in the protocols.
	Incomplete denaturation of target DNA	When dot or slot blotting, use the double denaturation methods described in *UNIT 2.9B*, or blot onto positively charged nylon.
	Blocking agents interfering with the target-probe interaction	If using a nylon membrane, leave the blocking agents out of the hybridization solution.
	Final wash was too stringent	Use a lower temperature or higher salt concentration. If necessary, estimate T_m as described in *UNIT 6.4*.
	Hybridization temperature too low with an RNA probe	Increase hybridization temperature to 65°C in the presence of formamide (see alternate protocol).
	Hybridization time too short	If using formamide with a DNA probe, increase the hybridization time to 24 hr.
	Inappropriate membrane	Check the target molecules are not too short to be retained efficiently by the membrane type (see Table 2.9.1).

continued

Table 2.10.4 Troubleshooting Guide for DNA Blotting and Hybridization Analysis[a], continued

Problem	Possible cause[b]	Solution
	Problems with electroblotting	Make sure no bubbles are trapped in the filter-paper stack. Soak the filter papers thoroughly in TBE before assembling the blot. Used uncharged rather than charged nylon.
Spotty background	Unincorporated nucleotides not removed from labeled probe	Follow protocols described in *UNIT 3.4*.
	Particles in the hybridization buffer	Filter the relevant solution(s).
	Agarose dried on the membrane	Rinse membrane in 2× SSC after blotting.
	Baking or UV cross-linking when membrane contains high salt	Rinse membrane in 2× SSC after blotting.
Patchy or generally high background	Insufficient blocking agents	See text for of discussion of extra/alternative blocking agents.
	Part of the membrane was allowed to dry out during hybridization or washing	Avoid by increasing the volume of solutions if necessary.
	Membranes adhered during hybridization or washing	Do not hybridize too many membranes at once (ten minigel blots for a hybridization tube, two for a bag is maximum).
	Bubbles in a hybridization bag	If using a bag, fill completely so there are no bubbles.
	Walls of hybridization bag collapsed on to membrane	Use a stiff plastic bag; increase volume of hybridization solution.
	Not enough wash solution	Increase volume of wash solution to 2 ml/10 cm^2 of membrane.
	Hybridization temperature too low with an RNA probe	Increase hybridization temperature to 65°C in the presence of formamide (see alternate protocol).
	Formamide needs to be deionized	Although commercial formamide is usually satisfactory, background may be reduced by deionizing immediately before use.
	Labeled probe molecules are too short	1. Use a ^{32}P-labeled probe as soon as possible after labeling, as radiolysis can result in fragmentation. 2. Reduce amount of DNase I used in nick translation (*UNIT 3.5*).
	Probe concentration too high	Check that the correct amount of DNA has been used in the labeling reaction.
	Inadequate prehybridization	Prehybridize for at least 3 hr with nitrocellulose or 15 min for nylon.
	Probe not denatured	Denature as described in the protocols.
	Inappropriate membrane type	If using a nonradiocative label, check that the membrane is compatible with the detection system.
	Hybridization with dextran sulfate	Dextran sulfate sometimes causes background hybridization. Place the membrane between Schleicher and Schuell no. 589 WH paper during hybridization, and increase volume of hybridization solution (including dextran sulfate) by 2.5%.

continued

Table 2.10.4 Troubleshooting Guide for DNA Blotting and Hybridization Analysis[a], continued

Problem	Possible cause[b]	Solution
	Not enough SDS in wash solutions	Check the solutions are made up correctly.
Extra bands	Final wash was not stringent enough	Use a higher temperature or lower salt concentration. If necessary, estimate T_m as described in *UNIT 6.4*.
	Probe contains nonspecific sequences (e.g., vector DNA)	Purify shortest fragment that contains the desired sequence.
	Target DNA is not completely restriction digested	Check the restriction digestion (*UNIT 3.1*).
	Formamide not used with an RNA probe	RNA-DNA hybrids are relatively strong but are destabilized if formamide is used in the hybridization solution.
Nonspecific background in one or more tracks	Probe is contaminated with genomic DNA	Check purification of probe DNA. The problem is more severe when probes are labeled by random printing. Change to nick translation.
	Insufficient blocking agents	See text for of discussion of extra/alternative blocking agents.
	Final wash did not approach the desired stringency	Use a higher temperature or lower salt concentration. If necessary, estimate T_m as described in *UNIT 6.4*.
	Probe too short	Sometimes a problem with probes labeled by random priming. Change to nick translation.
Cannot remove probe after hybridization	Membrane dried out after hybridization	Make sure the membrane is stored moist between hybridization and stripping.
Decrease in signal intensity when reprobed	Poor retention of target DNA during probe stripping	1. Check calibration of UV source if cross-linking on nylon. 2. Use a less harsh stripping method (support protocol).

[a]Based on Dyson (1991).

[b]Within each category, possible causes are listed in decreasing order of likelihood.

use. If a cDNA clone is used as the probe, or for the in vitro synthesis of an RNA probe, then blockage of sites with high affinity for poly(A)$^+$ sequences often reduces background. This is achieved by using 10 µg/ml of poly(A) DNA as the blocking agent.

Anticipated Results

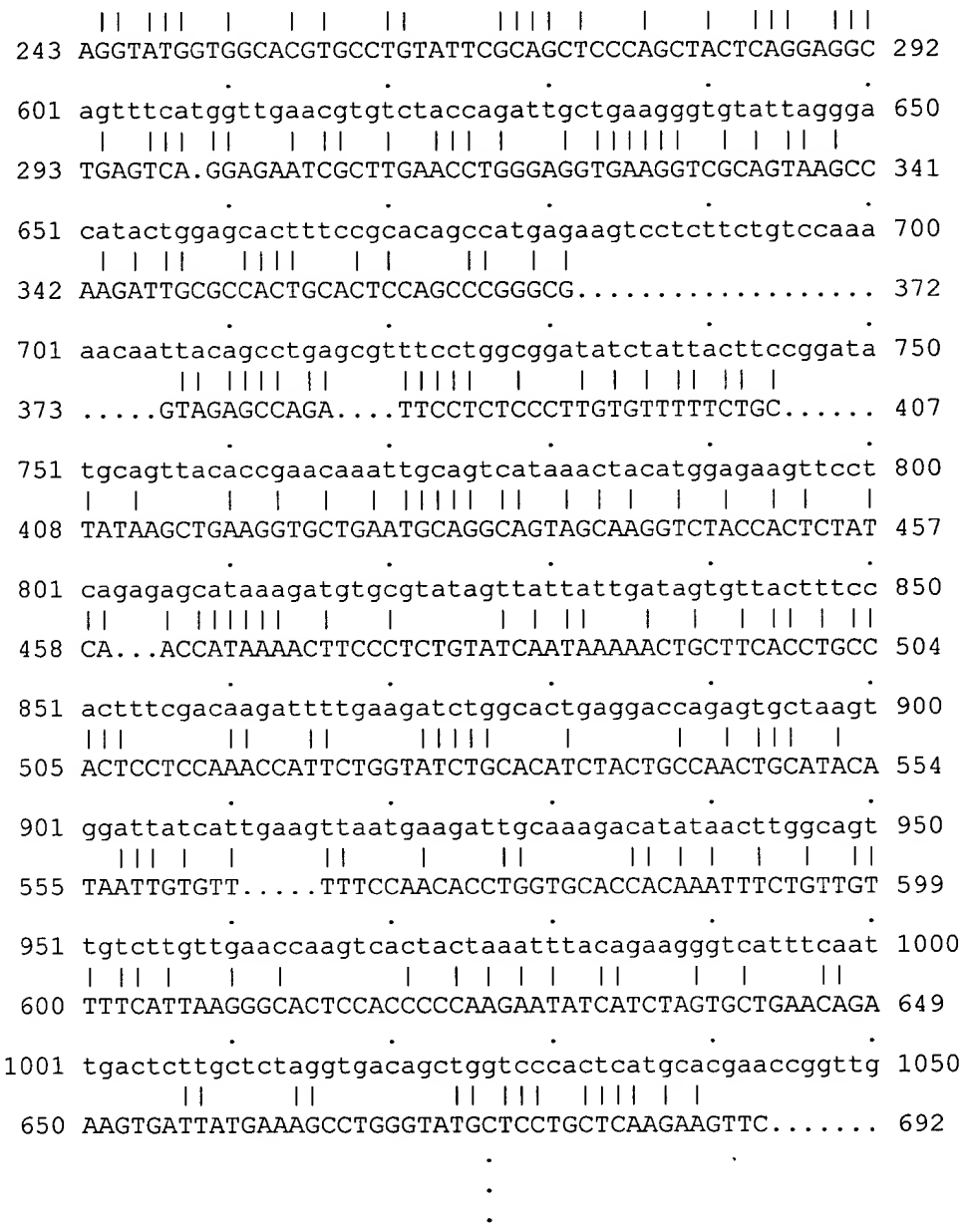
Using either a nitrocellulose or nylon membrane and a probe labeled to $\geq 10^8$ dpm/µg, there should be no difficulty in detecting 10 pg of a single copy sequence in human DNA after 24 hr autoradiography.

Time Considerations

The hybridization experiment can be completed in 24 hr, the bulk of this being taken up by the overnight incubation. Prehybridization takes 3 hr for a nitrocellulose membrane or 15 min for nylon. Post-hybridization washing to high stringency can usually be completed in 1.5 hr. If a single-copy sequence in human DNA is being probed, the hybridization step may be extended to 24 hr, with a concomitant increase in the length of the experiment as a whole.

The length of time needed for autoradiography depends on the abundance of the target sequences in the blotted DNA. Adequate exposure can take anything from overnight to several days.

```
 GAP of: 1107sid1  check: 4817  from: 1  to: 1474

WPDEF Case 1107 SEQ ID NO: 1
Case 1107 Rad51-like sequences.  From SEQ LISTING.

 to: 1107AI184177nt  check: 4275  from: 1  to: 692

WPDEF Case 1107 Rad51C GB AI184177
Case 1107 Rad51C GB AI184177 EST
AI184177. qf46e08.x1 Soares. . .[gi:3734815] Taxonomy, LinkOut
IDENTIFIERS
dbEST Id:       1947926
EST name:       qf46e08.x1 . . .

 Symbol comparison table: /app/gcg/10.2/gcgcore/data/rundata/nwsgapdna.cmp
 CompCheck: 8760

        Gap Weight:     50      Average Match: 10.000
     Length Weight:      3   Average Mismatch:  0.000

           Quality:   2335             Length:   1474
             Ratio:  3.374               Gaps:      7
 Percent Similarity: 40.751   Percent Identity: 40.751

       Match display thresholds for the alignment(s):
                   | = IDENTITY
                   : =   5
                   . =   1

 1107sid1 x 1107AI184177nt August 28, 2002 17:29  ..

                                    .
                                    .
                                    .
                .          .          .          .          .
    301 gcaatcaccatgggagatcaatctggctctagaaatggaccacaacagaa 350
                  | |||      |    |          ||  |
      1 ......TTTTTTGGATTAACAAGAGAAAGAGTTTATTAATTGTGCATAGT 44
                .          .          .          .          .
    351 gtacgtttcaggagcccagaatgcctgggatatgttctctgatgagctgt 400
        | | ||  ||  |     || | |   |  ||    |||  | |    ||
     45 GCATGTAACAACACAGGGGAGTCCAGAGATTAGTAACTCAAAAGTTTGGT 94
                .          .          .          .          .
    401 cacagaaacacatcactactggttctggtgacctcaatgacatacttggt 450
         | |     |  |    || |||| |      |  |||  ||    |  |
     95 TAGATTTGGAGCTGGGCACAGGTTTTCACACTCATAATCCCAGCACT..T 142
                .          .          .          .          .
    451 ggcgggattcactgcaaagaagttactgagatcggtggcgtcccaggggt 500
         | |      ||     || |     |||| || |  |  |   |   ||
    143 AGGGAAGCCGACGTAGGAGGATCACTTGAGGTCAGAAGTTTGAGACCAGT 192
                .          .          .          .          .
    501 tggtaaaactcaactggggattcaactagcaatcaatgtacaaatcccag 550
          |   |||   |        |       | |  |||  | ||||
    193 CTGGCCAACATGATGAAACTCTGTCTCTACTAAAAATACAAAAATTAGCC 242
                .          .          .          .          .
    551 tggaatgtggtggccttggtgggaaagcagtttatatagatacagagggc 600
```



GAP of: 1107sid1 check: 4817 from: 1 to: 1474

WPDEF Case 1107 SEQ ID NO: 1
Case 1107 Rad51-like sequences. From SEQ LISTING.

to: 1107AI184177nt.reverse check: 6897 from: 1 to: 692

REVERSE-COMPLEMENT of: 1107AI184177nt check: 4275 from: 1 to: 692
WPDEF Case 1107 Rad51C GB AI184177
Case 1107 Rad51C GB AI184177 EST
AI184177. qf46e08.x1 Soares. . .[gi:3734815] Taxonomy, LinkOut
IDENTIFIERS
dbEST Id: 1947926 . . .

Symbol comparison table: /app/gcg/10.2/gcgcore/data/rundata/nwsgapdna.cmp
CompCheck: 8760

Gap Weight:	50	Average Match:	10.000
Length Weight:	3	Average Mismatch:	0.000
Quality:	2879	Length:	1475
Ratio:	4.160	Gaps:	3
Percent Similarity:	44.139	Percent Identity:	44.139

Match display thresholds for the alignment(s):
| = IDENTITY
: = 5
. = 1

1107sid1 x 1107AI184177nt.reverse August 28, 2002 17:37 ..

```
                                   .
                                   .
                                   .
                .          .          .          .          .
    351 gtacgtttcaggagcccagaatgcctgggatatgttctctgatgagctgt 400
                                     ||  |  |     | ||||
      1 ............................GAACTTCTTGAGCAGGAGCATA 22
                .          .          .          .          .
    401 cacagaaacacatcactactggttctggtgacctcaatgacatacttggt 450
        | |||     ||| |  |||   | |   |  ||  |||| || |||||
     23 CCCAGGCTTTCATAATCACTTTCTGTTCAGCACTAGATGATATTCTTGGG 72
                .          .          .          .          .
    451 ggcgggattcactgcaaagaagttactgagatcggtggcgtcccaggggt 500
        || ||  | | ||  |   ||   || || ||  |||| |  ||||| ||
     73 GGTGGAGTGCCCTTAATGAAAACAACAGAAATTTGTGGTGCACCAGGTGT 122
                .          .          .          .          .
    501 tggtaaaactcaactggggattcaactagcaatcaatgtacaaatcccag 550
        ||| ||||| ||| |  | || ||  | ||| |  |||| || || ||||
    123 TGGAAAAACACAATTATGTATGCAGTTGGCAGTAGATGTGCAGATACCAG 172
                .          .          .          .          .
    551 tggaatgtggtggccttggtgggaaagcagtttatatagatacagagggc 600
            | ||| ||  | |  ||  ||||||||| ||| ||||||||||
    173 AATGGTTTGGAGGAGTGGCAGGTGAAGCAGTTTTTATTGATACAGAGGGA 222
                .          .          .          .          .
    601 agtttcatggttgaacgtgtctaccagattgctgaagggtgtattaggga 650
```